



3 0081 013 992 809

Science Library
VANDERBILT
UNIVERSITY LIBRARY
NASHVILLE, TENNESSEE



VUL-0267

QA
273.6
.N37
1980
v.5

47654

[illegible]

DEMCO NO. 38-29B

Statistical Distributions in Scientific Work

Volume 5 — Inferential Problems and Properties

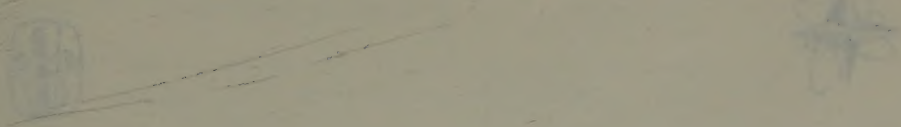
in Scientific Work

Volume 5 — Inferential Problems and Properties

The series is published by an international board of statisticians in cooperation with NATO Science Affairs Division
held at the University of London and the University of Cambridge
July 1, 1965 - July 1, 1966

- | | |
|---------------------------------------|--|
| A. Life Sciences | Plenum Publishing Corporation
London and New York |
| B. Physics | |
| C. Mathematical and Physical Sciences | D. Reidel Publishing Company
Dordrecht, Boston and London |
| D. Biological and Social Sciences | Springer-Verlag
Heidelberg, Berlin and New York |
| E. Applied Sciences | Alphen aan den Rijn and Dordrecht |

Printed in the Netherlands
Dordrecht, Holland
1966



D. Reidel Publishing Company
Statistical Distributions in Scientific Work
Volume 5 — Inferential Problems and Properties
Dordrecht, Holland
1966

NATO ADVANCED STUDY INSTITUTES SERIES

*Proceedings of the Advanced Study Institute Programme, which aims
at the dissemination of advanced knowledge and
the formation of contacts among scientists from different countries*

The series is published by an international board of publishers in conjunction
with NATO Scientific Affairs Division

A Life Sciences	Plenum Publishing Corporation London and New York
B Physics	
C Mathematical and Physical Sciences	D. Reidel Publishing Company Dordrecht, Boston and London
D Behavioural and Social Sciences	Sijthoff & Noordhoff International Publishers
E Applied Sciences	Alphen aan den Rijn and Germantown U.S.A.



Series C – Mathematical and Physical Sciences

Volume 79 – *Statistical Distributions in Scientific Work*

Volume 5 – *Inferential Problems and Properties*

Statistical Distributions in Scientific Work

Volume 5 – Inferential Problems and Properties

*Proceedings of the NATO Advanced Study Institute
held at the Università degli Studi di Trieste, Trieste, Italy,
July 10 – August 1, 1980*

edited by

CHARLES TAILLIE

Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, U.S.A.

GANAPATI P. PATIL

Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, U.S.A.

and

BRUNO A. BALDESSARI

*Istituto di Calcolo delle Probabilità, Facoltà di Scienze Statistiche Demografiche e Attuariali,
Università degli Studi di Roma, Italy*



D. Reidel Publishing Company

Dordrecht : Holland / Boston : U.S.A. / London : England

Published in cooperation with NATO Scientific Affairs Division

Library of Congress Cataloging in Publication Data

NATO Advanced Study Institute (1980: Trieste, Italy)
Statistical distributions in scientific work.

CIP

(NATO advanced study institutes series. Series C, Mathematical and physical sciences; v. 79)

Includes bibliographical references and indexes.

Contents: v. 4. Models, structures, and characterizations – v. 5. Inferential problems and properties – v. 6. Applications in physical, social, and life sciences.

1. Distribution (Probability theory)—Congresses. 2. Mathematical statistics—Congresses. I. Taillie, C. II. Patil, Ganapati P. III. Baldessari, Bruno. IV. North Atlantic Treaty Organization. Division of Scientific Affairs. V. Title. VI. Series.

QA273.6.N37 1980 519.5

81-12043

ISBN 90-277-1333-2 (v. 5)

AACR2

Published by D. Reidel Publishing Company
P.O. Box 17, 3300 AA Dordrecht, Holland

Sold and distributed in the U.S.A. and Canada
by Kluwer Boston Inc.,
190 Old Derby Street, Hingham, MA 02043, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, Holland

D. Reidel Publishing Company is a member of the Kluwer Group



All Rights Reserved

Copyright ©1981 by D. Reidel Publishing Company, Dordrecht, Holland
No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any informational storage and retrieval system, without written permission from the copyright owner

Printed in The Netherlands

TABLE OF CONTENTS

Foreword	xi
Program Acknowledgments	xiv
Reviewers of Manuscripts	xv
Contents of Edited Volumes	xvii
Preface	xxi

SECTION I: DISTRIBUTIONAL TESTING AND GOODNESS-OF-FIT

A Review of Distributional Testing Procedures and Development of a Censored Sample Distributional Test <i>Samuel S. Shapiro and Carlos Brain</i>	1
A Goodness-of-Fit Procedure for Testing Whether a Reliability Growth Model Fits Data that Show Improvement <i>Alan J. Gross and Samuel S. Shapiro</i>	25
Chi-Square Goodness-of-Fit Tests Based on Dependent Observations <i>Kamal C. Chanda</i>	35
An Asymptotically Distribution-Free Goodness-of-Fit Test for Families of Statistical Distributions Depending on Two Parameters <i>Fortunato Pesarin</i>	51
Conditionality Properties for the Bivariate Logarithmic Distribution with an Application to Goodness of Fit <i>A. W. Kemp</i>	57

SECTION II: PARAMETER ESTIMATION

A Survey of Estimating Distributional Parameters and Sample Sizes from Truncated Samples <i>Saul Blumenthal</i>	75
Parameter Estimation for Truncated Exponential Families <i>Bradford R. Crain and Loren Cobb</i>	87
Properties of the Maximum Likelihood Estimator of a Mixing Distribution <i>Bruce G. Lindsay</i>	95

Conditional Maximum Likelihood Estimation in Gaussian Mixtures <i>George E. Policello II</i>	111
Estimation of Pareto Parameters by Numerical Methods <i>W. Lyle Cook and Dean C. Mumme</i>	127
A New Estimation Procedure for the Three- Parameter Lognormal Distribution <i>Mario Chieppa and Pancrazio Amato</i>	133
SECTION III: HYPOTHESIS TESTING	
On the Asymptotic Distribution of the Multivariate Cramer-von Mises and Hoeffding-Blum-Kiefer- Rosenblatt Independence Criteria <i>Miklos Csörgö</i>	141
Complete Independence in the Multivariate Normal Distribution <i>Govind S. Mudholkar and Perla Subbaiah</i>	157
On Tests of Independence Under Bivariate Exponential Models <i>D. R. Weier and A. P. Basu</i>	169
On Tests for Detecting Change in the Multivariate Mean <i>M. S. Srivastava</i>	181
A Two-Dimensional t-Distribution and a New Test With Flexible Type I Error Control <i>G. Landenna and D. Marasini</i>	193
Testing Outliers in Multivariate Data <i>M. L. Tiku and M. Singh</i>	203
SECTION IV: APPROXIMATIONS	
A Normal Approximation for the Multivariate Likelihood Ratio Statistics <i>Govind S. Mudholkar and Madhusudan C. Trivedi</i>	219
Explicit Accurate Approximations for Fitting the Parameters of L_U <i>K. O. Bowman and L. R. Shenton</i>	231

A Method for the Evaluation of Cumulative Probabilities of Bivariate Distributions Using the Pearson Family <i>Rudolph S. Parrish and Rolf E. Bargmann</i>	241
---	-----

Transformation of a Discrete Distribution to Near Normality <i>Fabián Hernández and Richard A. Johnson</i>	259
---	-----

SECTION V: RELIABILITY AND LIFE TESTING

Multivariate Distributions in Reliability Theory and Life Testing <i>Henry W. Block and Thomas H. Savits</i>	271
---	-----

Hazard Rates Based on Isoprobability Contours <i>I. R. Goodman and Samuel Kotz</i>	289
---	-----

Failure Time Distributions: Estimates and Asymptotic Results <i>Janos Galambos</i>	309
---	-----

A Note on Shock Model Justification for IFR Distributions <i>Purushottam Laud and Roy Saunders</i>	319
---	-----

On the Mean Residual Life Function in Survival Studies <i>Ramesh C. Gupta</i>	327
--	-----

Identifiability Problems in the Theory of Competing and Complementary Risks - A Survey <i>Asit P. Basu</i>	335
---	-----

Dependence Concepts for Stochastic Processes <i>Dennis S. Friday</i>	349
---	-----

SECTION VI: MISCELLANEOUS

Some Distribution Theory Related to the Analysis of Subjective Performance in Inferential Tasks <i>J. Aitchison</i>	363
--	-----

Every Body Has Its Moments <i>E. G. Enns, P. F. Ehlers and S. Stuhr</i>	387
--	-----

Some Distributions in the Theory of Graphs <i>Michael Capobianco</i>	397
---	-----

Cograduation Between Statistical Distributions and Its Applications--A General Review · <i>Adolfo Gili</i>	415
<i>Author Index</i>	429
<i>Subject Index</i>	437

Statistical Distributions in Scientific Work

COMMITTEE

(Chairman: G. P. Patil)

Arnold, B.	Kemp, C. D.	Rao, C. R.
Baldessari, B. A.	Kotz, S.	Shapiro, S. S.
Cacoullos, T.	Mardia, K. V.	Stene, J.
Douglas, J. B.	Mosimann, J. E.	Taillie, C.
Engen, S.	Ord, J. K.	Tiago de Oliveira, J.
Folks, J. L.	Patil, G. P.	Warren, W. G.
Gross, A. J.	Ratnaparkhi, M. V.	Wise, M. E.

INTERNATIONAL SUMMER SCHOOL ON MODERN STATISTICAL DISTRIBUTION THEORY AND ITS APPLICATIONS (Trieste, Italy, 1980)

Director: B. A. Baldessari

Co-Director: C. Taillie

Host: L. Rondini

Secretary: B. Alles

Scientific Directors: S. Kotz, J. E. Mosimann, J. K. Ord, and G. P. Patil

Local Arrangements: E. Feoli, A. Kistoris (Secretary), S. Orviati, L. Rondini (Chairman),
M. Strassoldo, and M. Umani

SPONSORS

NATO Advanced Study Institutes Program

NATO Scientific Affairs Division

International Transfer of Science and Technology, Belgium

Department of Statistics, The Pennsylvania State University

Istituto di Calcolo delle Probabilità, Università di Roma

Istituto di Statistica, Università di Trieste

International Statistical Ecology Program

Office of Naval Research, USA

Consiglio Nazionale delle Ricerche, Italy

Regione Autonoma Friuli, Venezia Giulia, Italy

National Institutes of Health, USA

National Cancer Institute

National Heart, Lung, and Blood Institute

Fogarty International Center

Division of Computer Research and Technology

The Participants and Their Home Institutions and Organizations

PARTICIPANTS

- Ahmad, M., Saudi Arabia
 Aitchison, J., Hong Kong
 Akdeniz, F., Turkey
 Al-Ani, S., Iraq
 Alles, B., Pennsylvania
 Al-Zaid, A. H., England
 Ammassari, P., Italy
 Arnold, B. C., California
 Bajusz, B. A., Pennsylvania
 Baldessari, B. A., Italy
 Bargmann, R. E., Georgia
 Basu, A. P., Missouri
 Battaglia, F., Italy
 Beirlant, J., Belgium
 Blaesild, P., Denmark
 Block, H. W., Pennsylvania
 Blumenthal, S., Illinois
 Bochicchio, A., Italy
 Bologna, S., Italy
 Boswell, M. T., Pennsylvania
 Braumann, C. A., Portugal
 Brown, B. E., Massachusetts
 Cacoulios, T., Greece
 Canfield, R. V., Utah
 Capobianco, M., New York
 Chanda, K., Texas
 Charalambides, C. A., Greece
 Chieppa, M., Italy
 Chung, C. J. F., Canada
 Cobb, L., South Carolina
 Cocchi, D., Italy
 Cook, W. L., Idaho
 Csorgo, M., Canada
 Damiani, S., Italy
 De Lucia, L., Italy
 Dennis, B., Pennsylvania
 D'Esposito, M. R., Italy
 Diana, G., Italy
 Do Couto, H. T. Z., Brazil
 Dussauchoy, A., France
 Dyer, D. D., Texas
 Eicker, F., Germany
 Enns, E. G., Canada
 Feoli, E., Italy
 Ferrari, P., Italy
 Ferreri, C., Italy
 Finocchiaro, M., Italy
 Folks, J. L., Oklahoma
 Friday, D. S., Colorado
 Frishman, F., Maryland
 Galambos, J., Pennsylvania
 Gallo, F., Italy
 George, E. O., Nigeria
 Giavelli, G., Italy
 Gili, A., Italy
 Gomes, M. I. L., Portugal
 Gross, A. J., South Carolina
 Gupta, P. L., Maine
 Gupta, R. C., Maine
 Gyires, B., Hungary
 Hengeveld, R., Netherlands
 Hennemuth, R. C., Massachusetts
 Inal, H. C., Turkey
 Janardan, K. G., Illinois
 Jancey, R. C., Canada
 Kemp, A. W., England
 Kemp, C. D., England
 Kistoris, A., Italy
 Kotz, S., Maryland
 Landenna, G., Italy
 Langsaeter, T., Norway
 Laud, P. W., Illinois
 Lee, J. C., New Jersey
 Linder, E., Pennsylvania
 Lindsay, B., Pennsylvania
 Lukacs, E., Washington, D.C.
 Marasini, D., Italy
 Mardia, K. V., England
 Marvulli, R., Italy
 McDonald, J. B., Utah
 Mineo, Italy
 Miserocchi, M., Italy
 Morgan, P. R., Maryland
 Mosimann, J. E., Maryland
 Mudholkar, G. S., New York
 Norton, R. M., South Carolina
 Oksoy, D., Turkey
 Orviati, S., Italy
 Palmer, J. E., Massachusetts
 Panaretos, J., Ireland
 Papageorgiou, H., Greece
 Parrish, R. S., Georgia
 Patil, G. P., Pennsylvania
 Pesarin, F., Italy
 Plachky, D., West Germany
 Policello, G. E., Ohio
 Pollastri, A., Italy
 Poterasu, V., Romania
 Prato, G., Italy
 Provasi, C., Italy
 Ramachandran, G., England
 Ratnaparkhi, M. V., Maryland
 Reid, D. D., Australia
 Rigatti-Luchini, S., Italy
 Rondini, L., Italy
 Rossi, O., Italy
 Roux, J. J. J., South Africa
 Saunders, R., Illinois
 Sclove, S., Illinois
 Seshadri, V., Canada
 Shantaram, R., Michigan
 Shapiro, S. S., Florida
 Shimizu, R., Japan
 Singh, J., Pennsylvania
 Sobel, M., California
 Srivastava, M. S., Canada
 Srivastava, R. C., Ohio
 Stene, J., Denmark
 Stepanek, V., Czechoslovakia
 Strassoldo, M., Italy
 Taillie, C., Pennsylvania
 Tatlidil, H., Turkey
 Tiago de Oliveira, J., Portugal
 Tiku, M. L., Canada
 Tranquilli, G. B., Italy
 Tuncer, Y., Turkey
 Umani, M., Italy
 Uppuluri, V. R., Tennessee
 Vandemaale, M., Belgium
 Vedaldi, R., Italy
 Vik, G., Norway
 Villasenor, J. A., Mexico
 Weber, J. E., Arizona
 Wise, M. E., Netherlands
 Xekalaki, E., Ireland
 Zanni, R., Italy

AUTHORS NOT LISTED ABOVE

- Amato, P., Italy
 Barndorff-Nielsen, O., Denmark
 Becker, P. J., South Africa
 Bowman, K. O., Tennessee
 Brain, C., Florida
 Chen, T. L., California
 Crain, B. R., Oregon
 Davies, L., West Germany
 Davis, A. S., Wisconsin
 DeAngelis, R. J., Pennsylvania
 DuFrain, R. J., Tennessee
 Ehlers, P. F., Canada
 Ghosh, J. K., India
 Goodman, I. R., Washington, D.C.
 Hansen, E., California
 Hernandez, F., Wisconsin
 James, I., Australia
 Jensen, J. L., Denmark
 Johnson, R. A., Wisconsin
 Loukas, S., England
 Malley, J. D., Maryland
 Mathai, A. M., Canada
 Mazumder, B. S., India
 McNolty, F., California
 Midlarsky, M. I., Colorado
 Miller, M. C., III, South Carolina
 Mumme, D. C., Idaho
 Olsen, D. R., Texas
 Ord, J. K., England
 Rukhin, A. L., Indiana
 Savits, T. H., Pennsylvania
 Schaeffer, D. J., Illinois
 Shanmugam, R., Alabama
 Shenton, L. R., Georgia
 Singh, M., Canada
 Stulur, S., Canada
 Subbaiah, P., California
 Thomsen, W., West Germany
 Trivedi, M. C., New York
 Wenocur, R. S., Pennsylvania
 Weier, D. R., South Carolina

Foreword

The International Summer School on Statistical Distributions in Scientific Work was held in Trieste during July 1980 for a period of three weeks. The emphasis was on research, review, and exposition concerned with the interface between modern statistical distribution theory and real world problems and issues involving science, technology, and management. Both theory and applications received full attention at the School. The program consisted of a Short Intensive Preparation Course, a NATO Advanced Study Institute, and a Research Conference. While the relative composition of these activities varied somewhat in terms of instruction, exposition, research-review, research, and consultation, the basic spirit of each was essentially the same. Every participant was both a professor and a student.

The summer school was sponsored by the NATO Advanced Study Institutes Program; Consiglio Nazionale delle Ricerche, Italy; Regione Autonoma Friuli Venezia Giulia, Italy; National Institutes of Health, USA; Office of Naval Research, USA; The Pennsylvania State University; Universita di Roma; Universita di Trieste; International Statistical Ecology Program; International Transfer of Science and Technology, Belgium; and the participants and their home institutions and organizations.

Research papers, research-review expositions and instructional lectures were specially prepared for the program. These materials have been refereed and revised, and are now available in a series of several edited volumes and monographs.

BACKGROUND

It is now close to two decades since the International Symposium on Classical and Contagious Distributions was held in Montreal in 1963. It was the first attempt to identify the area of discrete distributions as a subject area by itself. The symposium was a great success in that it stimulated growth in the field and more importantly provided a certain direction to it. Next came the Biometric Society Symposium on Random Counts in Scientific Work at the annual meetings of the American Association for the Advancement of Science held in 1968. The first symposium had emphasized models and structures, the second one focused its attention on the useful role of discrete distributions in applied work.

Seven years ago, a Modern Course on Statistical Distributions in Scientific Work was held at the University of Calgary in 1974 under sponsorship of the NATO Scientific Affairs Division. The Program consisted of an Advanced Study Institute (ASI) followed by a Research Conference on Characterizations of Statistical Distributions. The purpose of the ASI was to provide an open forum with focus on different aspects of statistical distributions arising in scientific or statistical work. The purpose of the characterizations conference was to bring together research workers investigating characterization problems that have motivation in scientific concepts and formulations or that have application or potential use for statistical theory. The program was a great success. Participants still remember it very fondly for its scientific impact and its social and professional contact.

CALGARY PROGRAM

The edited Proceedings of the Calgary Program consist of three substantive volumes. They have been acknowledged to include a wealth of material ranging over a broad spectrum of the theory and applications of distributions and families of distributions. Most papers have been acknowledged for their content by reviewers in professional journals. The reviews have on the whole stressed the importance of these Proceedings as a successful effort to unify the field and to focus on main achievements in the area. Moreover, many of the papers which appeared in the Proceedings have been, and continue to be, quoted extensively in recent research publications. The Calgary Program of 1974 has had a definite and positive impact on stimulating further developments in the field of statistical distributions and their applications.

At the same time, essentially for economic reasons, the sciences, technology, and society are recognizing ever-expanding needs for quantification. The random quantities arising in conceptualization and modeling, in simulation, in data analysis, and in decision-making lead increasingly to various kinds of distributional problems and requests for solution. Statistical distributions remain an important and focal area of study. It is no surprise that the subject area of statistical distributions in scientific work is still advancing steadily.

Interestingly, the Calgary participants perceived this future need and concern. In anticipation, several prominent participants formed a Committee on Statistical Distributions in Scientific Work to discuss future plans and activities that would help consolidate and strengthen the subject area of statistical distributions and its applications on a continuing basis. The Committee identified the following needs and activities: (i) Preparation of a Comprehensive Dictionary and Bibliography of Statistical Distributions in Scientific Work, (ii) Preparation of Monographs and Modules on Important Distributions, Concepts, and Methods with Applications, and (iii) Planning and Organization of a Sequel to the Calgary Program.

DISTRIBUTIONAL ACTIVITIES

A well sustained seven year effort has produced a comprehensive three-volume set entitled *A Modern Dictionary and Bibliography of Statistical Distributions in Scientific Work*. The three volumes are: Volume 1, Discrete Models; Volume 2, Continuous Univariate Models; and Volume 3, Multivariate Models. The Dictionary covers several hundred distributional models and gives wherever possible their genesis, structural properties and parameters, random number generations, tabulations, graphs, and inter-relations through verbal statements as well as schematic diagrams. The Bibliography covers over ten thousand publications. Besides the usual reference information, each entry provides users listing (citation index), reviews, classification by distribution, inference and application, plus any special notes. The massive effort by the dictionary bibliography team consisting of M. T. Boswell, S. W. Joshi, G. P. Patil, M. V. Ratnaparkhi, and J. J. J. Roux needs to be specially acknowledged. So also the continuing interest and response of the professional community. It is hoped that the dictionary and bibliography effort will be a continuing activity serving the community with updated information from time to time.

On the monographs front, a lucid volume by J. B. Douglas, entitled *Analysis with Standard Contagious Distributions*, has been published. It should be of value to all those who are working with contagious distributions in one context or the other. More monographs are under preparation as follows:

Aitchison, J.: *Distributions on the Simplex of Their Applications*

Arnold, B. C.: *Pareto Distributions and Applications*

Cobb, L.: *Catastrophe Theory and Distributional Problems*

Folks, J. L. and Chhikara, R. S.: *Inverse Gaussian Distribution and Applications*

Mosimann, J. E.: *Analysis Using Size and Shape Variables*

Ord, J. K. and Patil, G. P.: *Introduction to Probability and Statistical Modeling*

Regarding the planning and organization of a sequel to the Calgary Program, the NATO Advanced Study Institutes Program encouraged part of the Committee to meet and assisted the Committee to have indepth discussions at Parma, Italy, in 1978. The following members were in attendance: B. A. Baldessari, T. Cacoullos, S. Engen, S. Kotz, J. E. Mosimann, J. K. Ord, G. P. Patil, C. Taillie, J. Tiago de Oliveira, W. G. Warren, and M. E. Wise. The intensive and open deliberations proved to be very constructive. The Committee felt unanimously that a follow-up to the Calgary ASI was very much needed, and that it should be held in 1980. Several institutions offered to host such an ASI. It was decided that the program be held in Italy. Bruno Baldessari and Livia Rondini assured the necessary support in this connection.

TRIESTE PROGRAM

A major purpose of the program was to give a unified and integrated view of different classes of distributions and to describe novel methodologies related to statistical distributions and/or their applications. Also, contributions on the description and characterization of distributions which are useful in a variety of fields of application were welcomed.

An application was prepared for the NATO ASI Program with G. P. Patil as the Chairman of the Organizing Committee, with B. Baldessari as the Director and C. Taillie as the Co-Director, with S. Kotz, J. E. Mosimann, J. K. Ord, and G. P. Patil as the Scientific Directors, and with L. Rondini as the Host. The NATO ASI program provided a positive response. Requests for the additional support needed were granted from within Italy and the USA. Participants and their institutions also extended a helping hand.

Spread over the three week period, the School had over 140 scientific participants and 50 accompanying persons from various countries around the world. The scientific program was more than full, and yet the overall program had a relaxing touch. Everything that the hosts, L. Rondini, A. Kostoris, S. Orviati, M. Strassoldo, M. Umani, and E. Feoli, did has been simply sweet and gratifying.

The Trieste program was a great success. Many have wondered as to when it would be again that they would meet and participate in another timely activity on statistical distributions in scientific work. If you have any thoughts or suggestions, please do not hesitate to let us know. I look forward to hearing from you.

Program Acknowledgments

For any program to be successful, mutual understanding and support among all participants are essential in directions ranging from critical to constructive and from cautious to constructive. The present program is grateful to the members of the Committee, and to the referees, advisors, sponsors and the participants for their timely advice and support.

Trieste is a beautiful place and so is the surrounding region. The Mediterranean around, the mountains nearby, and the campus on the top of a mountain provide a very scenic mosaic conducive for scholarship and communication. Italy has had a long tradition of research on distributional problems and related issues arising from uncertainty. It was only natural that the International Summer School on Statistical Distributions in Scientific Work met at Trieste.

The success of the program was due, in no small measure, to the endeavors of the Local Arrangements Committee. We thank L. Rondini, A. Kistoris, S. Orviati, M. Strassoldo, M. Umani, and E. Feoli for their hospitality and support.

And finally those who have assisted with the arduous task of preparing the materials for publication. Barbara Alles has been an ever cheerful and industrious secretary in the face of every adversity. Bonnie Burris, Bonnie Henninger, and Sandy Rothrock prepared the final versions of the manuscripts. Rani Venkataramani helped with the subject and author indexes. George Otto did the figures and artwork.

All of these nice people have done a fine job indeed. To all of them, our sincere thanks.

April 30, 1981

B. A. Baldessari
G. P. Patil
C. Taillie

Reviewers of Manuscripts

With appreciation and gratitude, the program acknowledges the valuable services of the following referees who have served as reviewers of manuscripts submitted to the program for possible publication. The editors thank the reviewers for their critical and constructive reviews.

M. Ahmad
University of Petroleum and Minerals
B. C. Arnold
University of California
A. C. Atkinson
Imperial College, London
L. J. Bain
University of Missouri
R. E. Bargmann
University of Georgia
O. Barndorff-Nielsen
Aarhus University
V. Barnett
University of Sheffield
A. K. Basu
Laurentian University
A. P. Basu
University of Missouri
L. V. Bellavista
University of Palermo
H. W. Block
University of Pittsburgh
S. Blumenthal
University of Illinois
M. T. Boswell
The Pennsylvania State University
L. A. Bruckner
Los Alamos Scientific Laboratory
R. V. Canfield
Utah State University
Ch. A. Charalambides
University of Athens
R. M. Cormack
University of St. Andrews
B. R. Crain
Portland State University
R. C. Dahiya
Old Dominion University
J. Darroch
Flinders University
A. P. Dawid
City University of London
B. Dennis
The Pennsylvania State University
P. J. Diggle
University of Newcastle upon Tyne
I. R. Dunsmore

University of Sheffield
A. Dussauchoy
Université Claude-Bernard
D. S. Friday
National Bureau of Standards
O. Frank
University of Lund
J. Galambos
Temple University
J. L. Gastwirth
George Washington University
D. V. Gokhale
University of California
M. I. Gomes
Faculty of Sciences of Lisbon
D. R. Grey
University of Sheffield
A. J. Gross
Medical University of South Carolina
A. M. Gross
Bell Laboratories
R. C. Gupta
University of Maine
D. M. Hawkins
CSIR, South Africa
R. Hengeveld
Catholic University, Nijmegen
T. Hettmansperger
The Pennsylvania State University
J. J. Higgins
University of South Florida
P. Holgate
Birkbeck College, London
H. K. Hsieh
University of Massachusetts
I. James
University of Western Australia
K. G. Janardan
Sangamon State University
N. L. Johnson
University of North Carolina
R. A. Johnson
University of Wisconsin
A. W. Kemp
University of Bradford
R. A. Kempton
Plant Breeding Institute, Cambridge

J. R. Kettenring
 Bell Laboratories
 C. G. Khatri
 Gujarat University
 S. Kotz
 University of Maryland
 P. R. Krishnaiah
 University of Pittsburgh
 I. J. Lauder
 University of Hong Kong
 J. C. Lee
 Wright State University
 H. J. Malik
 University of Guleph
 R. M. Marcus
 Equitable Life Assurance Society
 A. M. Mathai
 McGill University
 D. G. Morrison
 Columbia University
 N. E. Morton
 University of Hawaii
 J. E. Mosimann
 National Institutes of Health
 G. S. Mudholkar
 University of Rochester
 H. Nagao
 University of Tsukuba
 R. M. Norton
 College of Charleston
 J. K. Ord
 University of Warwick
 H. Papageorgiou
 University of Athens
 J. K. Patel
 University of Missouri
 P. K. Pathak
 University of New Mexico
 G. Policello
 Ohio State University
 B. Ramachandran
 Indian Statistical Institute
 G. Ramachandran
 Building Research Establishment
 M. V. Ratnaparkhi
 National Institutes of Health
 D. S. Robson
 Cornell University
 H. J. Rossberg
 Karl Marx University

J. J. J. Roux
 University of South Africa
 D. B. Rubin
 Educational Testing Service
 R. M. Schrader
 University of New Mexico
 A. K. Sen
 University of Illinois
 S. S. Shapiro
 Florida International University
 R. Shimizu
 Institute of Statistical Mathematics
 S. Shirahata
 Osaka University
 R. W. Shorrock
 Bell Canada
 R. Simon
 National Institutes of Health
 J. Singh
 Temple University
 M. D. Springer
 University of Arkansas
 M. S. Srivastava
 University of Toronto
 P. R. Tadikamalla
 University of Pittsburgh
 C. Taillie
 The Pennsylvania State University
 J. Taigo de Oliveira
 Faculty of Sciences of Lisbon
 G. L. Tietjen
 Los Alamos Scientific Laboratory
 R. C. Tripathi
 University of Texas
 A. A. Tsiatis
 St. Jude Children's Research Hospital
 G. G. Walter
 University of Wisconsin
 J. K. Wani
 University of Calgary
 W. G. Warren
 Forintek Canada Corp.
 J. E. Weber
 University of Arizona
 M. E. Wise
 Leiden University
 M. C. K. Yang
 University of Florida
 S. Zacks
 SUNY at Binghamton

Contents of Edited Volumes

Volume 4

MODELS, STRUCTURES, AND CHARACTERIZATIONS

455 pp.

Continuous Models: J. AITCHISON, Statistical Predictive Distributions. O. BARNDORFF-NIELSEN and P. BLAESILD, Hyperbolic Distributions and Ramifications: Contributions to Theory and Application. P. BLAESILD and J. L. JENSEN, Multivariate Distributions of Hyperbolic Type. L. COBB, The Multimodal Exponential Families of Statistical Catastrophe Theory. J. L. FOLKS and A. S. DAVIS, Regression Models for the Inverse Gaussian Distribution. V. SESHADRI, A Note on the Inverse Gaussian Distribution. V. R. R. UPPULURI, Some Properties of the Log-Laplace Distribution. J. J. J. ROUX and P. J. BECKER, Compound Distributions Relevant to Life Testing. I. JAMES, Distributions Associated with Neutrality Properties for Random Proportions. J. E. MOSIMANN and J. D. MALLEY, The Independence of Size and Shape Before and After Scale Change. J. AITCHISON, Distributions on the Simplex for the Analysis of Neutrality.

Discrete Models: E. XEKALAKI, Chance Mechanisms for the Univariate Generalized Waring Distribution and Related Characterizations. C. FERRERI, On a New Family of Discrete Distributions. R. SHANMUGAM and J. SINGH, On the Stirling Distribution of the First Kind. P. L. GUPTA and J. SINGH, On the Moments and Factorial Moments of a MPSD. T. CACOULLOS and H. PAPAGEORGIOU, On Bivariate Discrete Distributions Generated by Compounding. Ch. A. CHARALAMBIDES, Bivariate Generalized Discrete Distributions and Bipartitional Polynomials. M. AHMAD, A Bivariate Hyper-Poisson Distribution. B. GYIRES, On the Multinomial Distributions Generated by Stochastic Matrices and Applications.

Structural Properties: A. L. RUKHIN, Distributions with Sufficient Statistics for Multivariate Location Parameter and Transformation Parameter. E. LUKACS, Analytic Distribution Functions. S. L. SCLOVE, Some Recent Statistical Results for Infinitely Divisible Distributions. A. M. MATHAI, An Alternate Simpler Method of Evaluating the Multivariate Beta Function and an Inverse Laplace Transform Connected with Wishart Distribution. D. PLACHKY and W. THOMSEN, On a Theorem of Polya. J. C. LEE, Asymptotic Distributions of Functions of Eigenvalues.

Computer Generation: M. T. BOSWELL and R. J. DeANGELIS*, A Rejection Technique for the Generation of Random Variables with the Beta Distribution. C. D. KEMP and S. LOUKAS, Fast Methods for Generating Bivariate Discrete Random Variables. A. W. KEMP, Frugal Methods of Generating Bivariate Discrete Random Variables.

Characterizations: J. PANARETOS, A Characterization of the Negative Multinomial Distribution. R. C. GUPTA, On the Rao-Rubin Characterization of the Poisson Distribution. R. C. SRIVASTAVA, On Some Characterizations of the Geometric Distribution. M. V. RATNAPARKHI, On Splitting Model and Related Characterizations of Some Statistical Distributions. C. TAILLIE and G. P. PATIL, Rao-Rubin Condition for a Certain Class of Continuous Damage Models. J. J. J. ROUX and M. V. RATNAPARKHI, On Matrix-Variate Beta Type I Distribution and Related Characterization of Wishart Distribution. J. PANARETOS, On the Relationship Between the Conditional and Unconditional Distribution of a Random Variable. M. V. RATNAPARKHI, Some Bivariate Distributions of (X, Y) where the Conditional Distribution of Y , Given X , is Either Beta or Unit-Gamma. E. O. GEORGE and G. S. MUDHOLKAR, Some Relationships Between the Logistic and the Exponential Distributions. R. C. SRIVASTAVA, Some Characterizations of the Exponential Distribution Based on Record Values. C. TAILLIE, A Note on Srivastava's Characterization of the Exponential Distribution Based on Record Values. R. SHANTARAM, On the Stochastic Equation $X + Y = XY$. R. SHIMIZU and L. DAVIES, On the Stability of Characterizations of Non-Normal Stable Distributions.

Volume 5

INFERENCE PROBLEMS AND PROPERTIES

439 pp.

Distributional Testing and Goodness-of-Fit: S. S. SHAPIRO and C. BRAIN, A Review of Distributional Testing Procedures and Development of a Censored Sample Distributional Test. A. J. GROSS and S. S. SHAPIRO, A Goodness-of-Fit Procedure for Testing Whether a Reliability Growth Model Fits Data that Show Improvement. K. C. CHANDRA, Chi-Square Goodness-of-Fit Tests Based on Dependent Observations. F. PESARIN, An Asymptotically Distribution-Free Goodness-of-Fit Test for Families of Statistical Distributions Depending on Two Parameters. A. W. KEMP, Conditionality Properties for the Bivariate Logarithmic Distribution with an Application to Goodness of Fit.

Parameter Estimation: S. BLUMENTHAL, A Survey of Estimating Distributional Parameters and Sample Sizes from Truncated Samples. B. R. CRAIN and L. COBB, Parameter Estimation for Truncated Exponential Families. B. G. LINDSAY, Properties of the Maximum Likelihood Estimator of a Mixing Distribution. G. E. POLICELLO II, Conditional Maximum Likelihood Estimation in Gaussian Mixtures. W. L. COOK and D. C. MUMME, Estimation of Pareto Parameters by Numerical Methods. M. CHIEPPA and P. AMATO, A New Estimation Procedure for the Three-Parameter Lognormal Distribution.

Hypothesis Testing: M. CSORGO, On the Asymptotic Distribution of the Multivariate Cramer-von Mises and Hoeffding-Blum-Kiefer-Rosenblatt Independence Criteria. G. S. MUDHOLKAR and P. SUBBAIAH, Complete Independence in the Multivariate Normal Distribution. D. R. WEIER and A. P. BASU, On Tests of Independence Under Bivariate Exponential Models. M. S. SRIVASTAVA, On Tests for Detecting Change in the Multivariate Mean. G. LANDENNA and D. MARASINI, A Two-Dimensional t-Distribution and a New Test with Flexible Type I Error Control. M. L. TIKU and M. SINGH, Testing Outliers in Multivariate Data.

Approximations: G. S. MUDHOLKAR and M. C. TRIVEDI, A Normal Approximation for the Multivariate Likelihood Ratio Statistics. K. O. BOWMAN and L. R. SHENTON, Explicit Accurate Approximations for Fitting the Parameters of LU. R. S. PARRISH and R. E. BARGMANN, A Method for the Evaluation of Cumulative Probabilities of Bivariate Distributions Using the Pearson Family. F. HERMANDEZ and R. A. JOHNSON, Transformation of a Discrete Distribution to Near Normality.

Reliability and Life Testing: H. W. BLOCK and T. H. SAVITS, Multivariate Distributions in Reliability Theory and Life Testing. I. R. GOODMAN and S. KOTZ, Hazard Rates Based on Isoprobability Contours. J. GALAMBOS, Failure Time Distributions: Estimates and Asymptotic Results. P. LAUD and R. SAUNDERS, A Note on Shock Model Justification for IFR Distributions. R. C. GUPTA, On the Mean Residual Life Function in Survival Studies. A. P. BASU, Identifiability Problems in the Theory of Competing and Complementary Risks — A Survey. D. S. FRIDAY, Dependence Concepts for Stochastic Processes.

Miscellaneous: J. AITCHISON, Some Distribution Theory Related to the Analysis of Subjective Performance in Inferential Tasks. E. G. ENNS, P. F. EHLERS, and S. STUHR, Every Body Has Its Moments. M. CAPOBIANCO, Some Distributions in the Theory of Graphs. A. GILI, Cogradsation Between Statistical Distributions and Its Applications — A General Review.

Volume 6

APPLICATIONS IN PHYSICAL, SOCIAL, AND LIFE SCIENCES

445 pp.

Applications in the Physical Sciences: K. V. MARDIA, Recent Directional Distributions with Applications. J. K. GHOSH and B. S. MAZUMDER, Size Distribution of Suspended Particles-Unimodality, Symmetry and Lognormality. D. DYER, Offshore Oil/Gas Lease Bidding and the Weibull Distribution. F. McNOLTY and E. HANSEN, Statistical Distributions Occurring in Photoelectron Phenomena, Radar and Infrared Applications. K. G. JANARDAN and D. J. SCHAEFFER, Application of Discrete Distributions for Estimating the Number of Organic Compounds in Water. R. SHANMUGAM and J. SINGH, Some Bivariate Probability Models Applicable to Traffic Accidents and Fatalities. V. STEPANEK, Role and Use of Statistical Distributions in Information Theory as Applied to Chemical Analysis.

Applications in the Social Sciences: S. L. SCLOVE, Modeling the Distribution of Fingerprint Characteristics. M. I. MIDLARSKY, Stochastic Modeling in Political Science Research. P. R. MORGAN, Statistical Distribution Models in the Behavioral Sciences: A Review of Theory and Applications. J. B. McDONALD, Some Issues Associated with the Measurement of Income Inequality. C. TAILLIE, Lorenz Ordering Within the Generalized Gamma Family of Income Distributions. J. K. ORD, G. P. PATIL and C. TAILLIE, The Choice of a Distribution to Describe Personal Incomes. J. K. ORD, G. P. PATIL and C. TAILLIE, Relationships Between Income Distributions for Individuals and for Households.

Applications in the Life Sciences: M. E. WISE, Spike Interval Distributions for Neurons and Random Walks with Drift to a Fluctuating Threshold. J. STENE, Probability Distributions Arising from the Ascertainment and the Analysis of Data on Human Families and Other Groups. K. G. JANARDAN, D. J. SCHAEFFER and R. J. DuFRAIN, A Stochastic Model for the Study of the Distribution of Chromosome

Aberrations in Human and Animal Cells Exposed to Radiation or Chemicals. D. COCCHI, A Model for the Analysis of Platelet Survival. B. DENNIS, Extinction and Waiting Times in Birth-Death Processes: Applications to Endangered Species and Insect Pest Control. D. D. REID, The Poisson Lognormal Distribution and Its Use as a Model of Plankton Aggregation. A. J. GROSS and M. C. MILLER III, Some Applications of Statistical Distribution Theory to Biology and Medicine.

Extreme Values and Order Statistics: R. V. CANFIELD, D. R. OLSEN, and T. L. CHEN, Extreme Value Theory with Application to Hydrology. G. RAMACHANDRAN, Properties of Extreme Order Statistics and Their Application to Fire Losses and Earthquake Magnitudes. J. TIAGO DE OLIVEIRA, Statistical Choice of Univariate Extreme Models. M. I. GOMES, An i -Dimensional Limiting Distribution Function of Largest Values and Its Relevance to the Statistical Theory of Extremes. R. S. WENOCUR, Waiting Times and Return Periods to Exceed the Maximum of a Previous Sample. R. S. WENOCUR, Waiting Times and Return Periods Related to Order Statistics: An Application of Urn Models.

Essentially because of the present economic conditions, the sciences, technology, and society are recognizing ever-expanding needs for quantification. The random quantities arising in conceptualization and modeling, in simulation, in data analysis, and in decision making lead increasingly to various kinds of distributional problems and requests for solution. Statistical distributions remain an important and focal area of study.

Preface

These three volumes constitute the edited Proceedings of the NATO Advanced Study Institute on Statistical Distribution Theory and its Applications held at the University of Trieste from July 10-August 1, 1980. The general title of the volume is *Statistical Distributions in Scientific Work*, a continuation from the Proceedings of an earlier program held at the University of Calgary during the summer of 1974, which brought out volumes 1, 2, and 3. The present volumes are: Volume 4 — Models, Structures, and Characterizations; Volume 5 — Inferential Problems and Properties; and Volume 6 — Applications in Physical, Social, and Life Sciences. These are based on the research-review expositions, instructional lectures, and research papers specially prepared for the program by the invited researchers and expositors.

The planned activities of the Institute consisted of lucid perceptive lectures and expositions, seminar lectures, study group discussions, tutorials, and individual study. The activities included meetings of editorial committees to discuss editorial matters for these proceedings which consist of the contributions that have gone through the usual refereeing process. The overall perspective of the program is provided by the Chairman of the Organizing Committee, Professor G. P. Patil, in his Foreword to the Volumes as summarized from his inaugural address to the Institute.

The Proceedings are being published in three volumes. All together, they consist of 15 topical sections of 100 contributions of 1260 pages of research, review, and exposition. Subject and author indexes also appear at the end of each volume. Effort has been made to keep the title and the content of each volume mutually consistent. However, it is quite possible that a different composition would have looked equally natural!

We view this program as a continuation of the tradition established by the pioneering 1963 Montreal Symposium which identified and consolidated statistical distributions as a separate field of statistical inquiry. The tradition was further carried on and amplified by the 1974 Calgary program. It was reassuring to see several participants at Trieste that were present at Montreal and/or Calgary. A number of new and young faces were also visible at Trieste. The papers in these Proceedings should reflect the recent and current developments and mirror the growth and maturity of the discipline and its integration within the general framework of applied statistics and related quantitative studies.

While working in the field of statistical distributions in general, it is often tempting to tackle isolated problems involving formal generalizations. One at times loses sight of the underlying probabilistic model even in this process. While this generalization approach may be quite acceptable from the mathematical point of view, it does however result, on occasion, in statistically unjustified theoretical exercises. There has been some justified criticism voiced by practitioners that we are losing touch with reality. A purpose of the Trieste program was to help generate a constructive dialogue between theory and application.

The program covered a broad spectrum of topics. Models and structures theme touched base with continuous models, discrete models, properties, computer generation, and characterizations. Inferential problems and properties included distribu-

tional testing and goodness-of-fit, parameter estimation, hypothesis testing, approximations, reliability and life testing. Real world problems were drawn from the physical sciences, social sciences and life sciences, and also included work on extreme values and order statistics. Thus, the formal and informal dialogues provided a panorama of the distributional field both in theory and in application. These published volumes constitute an effort to share those Proceedings with the interested reader. The spark and the spontaneity of a lively dialogue do not necessarily transmit themselves through written proceedings. We hope and trust, however, that the reader will instead reap the benefit from the careful preparation and editing through which each paper has gone.

In any collaborative effort of this magnitude and nature, the enthusiastic support of a large number of individuals and institutions is a prerequisite for success. We are extremely grateful to all of our sponsors, participants, and the hosts. Also to our ever-cheerful program secretary, Barbara Alles, who has managed to keep the program moving in every sense of the word.

These three volumes have been included in the ongoing NATO Advanced Study Institutes Series. They are published by the D. Reidel Publishing Company, a member of the Board of Publishers of the NATO ASI Series. It is only proper that we conclude here with our sincere thanks to both the Publisher and the NATO Scientific Affairs Division for these co-operative arrangements.

April 30, 1981

Charles Taillie
Ganapati P. Patil
Bruno A. Baldessari

A REVIEW OF DISTRIBUTIONAL TESTING PROCEDURES AND DEVELOPMENT OF A CENSORED SAMPLE DISTRIBUTIONAL TEST

**SAMUEL S. SHAPIRO and
CARLOS W. BRAIN**

Department of Mathematical Sciences
Florida International University
Miami, Florida 33199 USA

SUMMARY. A review of recent procedures for tests of distributional assumptions is given. Test procedures are grouped into three categories: regression type tests, probability integral transformation tests and special feature tests. The application of regression tests to distributions with location and shape parameters with emphasis on the normal, multivariate normal, exponential, and Weibull distributions and the advantages and limitations of recently improved EDF procedures are discussed. Tests which use characteristics unique to each of the normal, exponential, gamma, extreme value and Weibull distributions are also discussed.

New regression test procedures for censored samples are discussed in general and then applied to the exponential and normal distributions.

KEY WORDS. distributional tests, W tests, EDF tests, censored samples.

1. INTRODUCTION

The subject of testing for distributional assumptions dates back to the work of K. Pearson (1900) in which he devised the chi-square goodness of fit test. Statisticians have been interested in this subject because inferences based on an assumed statistical model can be quite poor if the assumption is incorrect. In the last ten years there has been a number of new procedures proposed. The first part of this paper represents an attempt to group these procedures into general categories and present a

review of the latest developments. The review is limited to recent work and no attempt has been made to include every procedure.

The second part of the paper develops a rationale for a test procedure which can be used with censored samples. This idea is then applied to develop test statistics for the exponential and normal distributions.

2. REVIEW

Tests for distributional procedures have been categorized into three groupings: regression type tests, probability transformation tests and special feature tests. The latter category contains procedures which use a special characteristic of the null distribution as the test criterion, e.g., the standardized third moment, $\sqrt{b_1}$, test for the normal distribution makes use of the fact that $\sqrt{\beta_1} = 0$. The advantage of the tests in the first two categories is that they can be generalized for many null distributions while the special feature tests cannot. In passing it should be noted that the chi-square goodness of fit test has been omitted. Many practitioners have found its power properties lacking or dislike the necessity of discretizing the data.

2.1 Regression Tests. The best known of the regression procedures is a subjective, graphical procedure known as probability plotting. This procedure is available for many distributions and can be also used with censored samples. A probability plot can be considered as the graphical representation of the regression of the ordered observations on the expected values of the order statistics from a standardized distribution (distribution with unit scale parameter and zero location parameter). Thus,

$$y_{i/n} = \mu + \sigma m_{i/n} + e_i \quad (1)$$

represents such a model, where

$y_{i/n}$ is the i th ordered observation in a sample of size n ,

$m_{i/n}$ is the corresponding expected value of the i th order statistic from a standardized distribution

μ is the location parameter,

σ is the scale parameter, and

e_i is the measurement error.

(Subsequently the subscript n above will be suppressed and in the following y_i represents the i th ordered observation.) In practice one does not need to know the m_i 's. Blom (1958) showed that $m_i \approx F^{-1}(\Pi_i)$ where

$$\Pi_i = (i - \alpha_i) / (n - \alpha_i - \beta_i + 1)$$

and F^{-1} is the inverse of the standardized distribution function. Thus, a plot of y_i vs m_i can be carried out by using specially scaled paper and plotting y_i vs Π_i ; the scaling converts Π_i to $F^{-1}(\Pi_i)$. The values of α_i and β_i vary from distribution to distribution; however, a commonly used compromise is to set $\alpha_i = \beta_i = 1/2$. The rationale behind the technique is that if the y_i 's are a random sample from the null distribution represented by the m_i 's then a straight line plot, up to random fluctuations, should result. Departures from linearity are indicative of lack of fit. The procedure of course is subjective and there is no type I error associated with the procedure.

Detailed descriptions of how to make and use probability plots can be found in Shapiro (1980) and Hahn and Shapiro (1967).

In order to make this procedure more objective, Shapiro and Wilk (1965) suggested testing for linearity by considering b , the slope of the regression line, which provides an estimate of σ in (1). This estimate of σ , when squared, is then compared to S^2 , the total sum of squares about the mean, which is an estimate of $(n-1)\sigma^2$ much the same way as is done in an analysis of variance procedure. If the model is correct, hence, the straight line is the proper model, then b^2 provides an estimate of σ^2 ; however, if the model is inappropriate then b^2 will not be an appropriate estimator. The estimator S^2 always provides an estimate of $(n-1)\sigma^2$. This ratio is independent of location and scale parameters and thus does not require knowing their values.

Using this rationale Shapiro and Wilk developed tests for the normal (1965) and the two-parameter exponential (1972) distributions. The test statistics were respectively

$$W_N = \left[\sum_{i=1}^{[n/2]} a_{n-i+1} (y_{n-i+1} - y_i) \right]^2 / \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2)$$

and
$$W_E = \frac{n(\bar{y} - y_1)^2}{(n-1) \sum_{i=1}^n (y_i - \bar{y})^2}$$

where $[n/2]$ is the greatest integer symbol and the a_i 's are constants derived from the expected values and covariance matrix of the normal order statistics and are required for linear estimation of σ based on the slope of the regression line. The authors provided the a_i 's and percentiles of W_N for samples up to size 50 and the percentiles of W_E up to sample size 100. Shapiro and Francia (1972) extended the normality test up to sample size 100 by replacing the a_i 's with coefficients that depend only on the expected values of the order statistics. The modified procedure

$$W'_N = \left[\sum_{j=1}^{[n/2]} b_j (y_{n-i+1} - y_i) \right]^2 / \sum (y_i - \bar{y})^2$$

can be extended to even higher sample sizes since the b_i 's are simple functions of the m_i 's. Shapiro (1980) provided a table of b_i values and percentiles for n up to 99.

Stephens (1978) provided a modified form of the W_E procedure for the case where the location parameter is known. In this procedure the known location parameter, μ (the minimum possible value) is included as an observation. This increases the sample size to $n+1$ and the W_E statistic is computed with μ replacing y_1 in the formula and $n+1$ replacing n . The same table of percentiles is used except that sample size is considered to be $n+1$.

Sarkadi (1975) proved that the W_N and W'_N are asymptotically equivalent and that both are consistent. Weisberg (1974) showed that the percentage points for W_N and W'_N were similar and W_N values could be used to estimate the W'_N values for $n \leq 50$.

The power of the W procedure has been studied by many authors who have generally used it as the standard to which a new procedure is compared. In most of these cases the power of the W procedure ranks high for wide ranges of differing shaped distributions.

Another approach to assessing the linearity of (1) is to fit higher order polynomials to the model and test for lack of fit. This was first suggested by Shapiro (1964) and developed by LaBrecque (1977). In this latter paper three alternative models were used. The first was

$$y_i = \mu + \sigma m_i + \alpha \phi_2(m_i) + \beta \phi_3(m_i) + e_i$$

and the second and third set $\alpha = 0$ and then $\beta = 0$, respectively. The $\phi_j(m_i)$ are orthogonal polynomials of order j and a test of the hypothesis $\alpha = 0$ and/or $\beta = 0$ (depending on the model) provides a technique for testing for goodness of fit. LaBrecque suggested in testing for normality use of the quadratic function against skewed alternatives and the cubic function against symmetric alternatives. The three tests suggested for normality were

$$F_1 = [\hat{a}^2 + \hat{b}^2]/2s^2, F_2 = \hat{a}^2/s^2, F_3 = \hat{b}^2/s^2$$

$$\text{where } \hat{a} = \begin{cases} \sum_{i=1}^{n/2} a_i (y_i - y_{n-i+1}) & \text{for } n \text{ even} \\ \sum_{i=1}^{[n/2]} a_i (y_i - y_{n-i+1}) + a_{\frac{n+1}{2}} \frac{y_{\frac{n+1}{2}}}{2} & \text{for } n \text{ odd} \end{cases}$$

$$\hat{b} = \sum_{i=1}^{[n/2]} b_i (y_i - y_{n-i+1}), \quad s^2 = \sum (y_i - \bar{y})^2 / (n-1)$$

and the a_i 's and b_i 's for $n \leq 64$ were provided in the paper. LaBrecque provided a table of percentiles for $n \leq 12$ and a simple function to compute the percentiles for $n > 12$.

Another version of the regression tests was suggested by Filliben (1975). In this procedure the probability plot was considered as a regression of the ordered observations on the medians of the order statistics (m'_i) and the author used the coefficient of correlation between the y 's and the (m') 's as the test criterion. Thus,

$$r = \frac{\sum (y_i - \bar{y}) (m'_i - \bar{m}')}{\sqrt{\sum (y_i - \bar{y})^2 \sum (m'_i - \bar{m}')^2}}$$

is used as the test statistic. The rationale here is that if equation (1) holds, i.e., a straight line is the appropriate fit, then r will be close to one. One should note that r will always be positive and relatively high since the y_i 's and the (m'_i) 's are both ordered. The distribution of r depends only on the sample size and the null distribution and Filliben provided the percentiles of r up to $n = 100$ for the normal distribution. The author also recommended obtaining the medians as follows. The normal order statistic medians, (m'_i) 's, are related to the uniform order statistic medians, g_i 's, through the inverse probability integral transformation. An approximation to the uniform order statistic medians is

$$g_n = 1 - g_1 = (.5)^{1/n} \text{ and } g_i = (1 - .3175)/(n + .365)$$

$$i = 2, 3, \dots, n-1 \text{ and } m'_i = F^{-1}(g_i).$$

The calculation of r can be simplified for the normal distribution since $\bar{m}' = 0$ and $m'_{n-i+1} = -m'_i$; thus

$$r = \frac{\sum_{i=1}^n m'_i y_i}{\sqrt{\sum_{i=1}^n m_i'^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where $y_1 < y_2 < \dots < y_n$.

The regression procedures are quite general and the test rationale can be used for any location-scale parameter family or one that can be transformed to such a family. However, since the coefficients and percentiles depend on the null hypothesis these must be generated for each case. For example, Shapiro (1964) derived the W test for the uniform distribution as

$$W_u = (y_n - y_1)^2 / \sum (y_i - \bar{y})^2$$

and Brain and Shapiro (1980) developed a statistic for testing for the Weibull distribution. This latter procedure makes use of the fact that the function $x_i = \ln y_i$ transforms a Weibull variate to a type 1 extreme value (smallest value) distribution which is a location-scale parameter family. The test statistic is

$$W_w = [.6079 L_2 - .2570 L_1]^2 / n^2 \sum (x_i - \bar{x})^2$$

where $x_i = \ln y_i$, $L_1 = \sum_{i=1}^n a_i x_i$, $L_2 = \sum_{i=1}^n b_i x_i$

$$a_i = \ln \left(\frac{n+1}{n+1-i} \right), \quad b_i = a_i (1 + \ln a_i) - 1, \quad i = 1, 2, \dots, n-1$$

$$a_n = n - \sum_{i=1}^{n-1} a_i, \quad b_n = .4228n - \sum_{i=1}^{n-1} b_i.$$

The numerator of this statistic uses a linear estimator of the scale parameter (for extreme value distribution) suggested by D'Agostino (1971). Brain and Shapiro (1980) provide the necessary constants and percentiles for sample sizes up to 50.

A test of multivariate normality using the W_N procedure was developed by Malkovich and Afifi (1973) using S. N. Roy's (1957) "union-intersection" principle of test construction. The test is constructed as follows. If \tilde{Y} is a multivariate normal random variable of dimension $p \times 1$ then $\tilde{C}'\tilde{Y}$ is univariate normal for all constant vectors \tilde{C} , $\tilde{C} \neq 0$. A multivariate normal sample $\tilde{Y}_1, \dots, \tilde{Y}_n$ can then be reduced to a univariate normal sample $\tilde{C}'\tilde{Y}_1, \dots, \tilde{C}'\tilde{Y}_n$, and the order statistics of this univariate sample can be used to evaluate the W_N statistic given in (2). Thus, if one can find the vector \tilde{C}^* that minimizes W_N and use the corresponding values of $\tilde{C}^*\tilde{Y}$ in W_N then a test for normality is given by

$$W_N(\tilde{C}^*) = \min_{\tilde{C}} W_N(\tilde{C}) \leq K_w$$

where K_w is a constant. A solution for \tilde{C}^* does not exist but an approximation is given by the authors. The test is carried out as follows:

1. Let \tilde{Y}_m be the observation vector for which

$$(\tilde{Y}_m - \bar{\tilde{Y}})' \tilde{A}^{-1} (\tilde{Y}_m - \bar{\tilde{Y}}) = \max_{1 \leq j \leq n} (\tilde{Y}_j - \bar{\tilde{Y}})' \tilde{A}^{-1} (\tilde{Y}_j - \bar{\tilde{Y}})$$

where $\tilde{A} = \sum_{j=1}^n (\tilde{Y}_j - \bar{\tilde{Y}}) (\tilde{Y}_j - \bar{\tilde{Y}})'$, and $\bar{\tilde{Y}}$ is the sample mean vector.

2. Order the statistics

$$u_{(j)} = (\mathbf{Y}_{\sim m} - \bar{\mathbf{Y}})' \mathbf{A}^{-1} (\mathbf{Y}_j - \bar{\mathbf{Y}}), \quad j = 1, 2, \dots, n$$

and denote them by u_1, u_2, \dots, u_n .

3. The test statistic is

$$W_{MN}^* = \frac{\left(\sum_{i=1}^{[n/2]} a_{n-i+1} (u_{n-i+1} - u_i) \right)^2}{(\mathbf{Y}_{\sim m} - \bar{\mathbf{Y}})' \mathbf{A}^{-1} (\mathbf{Y}_{\sim m} - \bar{\mathbf{Y}})}$$

where the a_i 's are given in Shapiro and Wilk (1965).

In summary, the regression tests are quite general, can be generated for most location-scale parameter families for which linear estimation of the scale parameter can be accomplished, are scale and location invariant and in the studies to date Shapiro, Wilk and Chen (1968), LaBrecque (1977) and Filliben (1975) have been shown to have good power against a wide range of alternatives relative to other procedures. The power differentials among the various regression procedures are small and vary with sample size and alternative distribution. Choice of any particular one could be dependent on ease of computation or availability of a computer program.

2.2 Probability Integral Transformation Tests. It is a well-known fact that the distribution function is uniformly distributed. The transformation

$$p_i = \int_{-\infty}^{y_i} f(\tau) d\tau$$

converts the random variable y_i with density $f(y)$ to p_i which has a uniform density. Thus, a test for a distributional assumption can be carried out by transforming the observations y_i to p_i and testing p_i 's for uniformity. The transformation can be performed if the parameters of $f(y)$ are known. A series of tests have been developed using this rationale. The best known of these are the Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling and Watson procedures. In addition, there have been a number of suggestions made in an attempt to improve the power of these procedures, for example, Durbin (1961) suggested a series of transformations which in the non-null case would tend to make the transformed variates more non-uniform. The major advantage of these procedures, often called Empirical

Distribution Function or EDF procedures, is that if the parameters are known the distribution of the test statistic does not depend on the null distribution and unlike the regression procedures, one set of percentiles work for all null distributions. The major disadvantages of these procedures are that when the parameters are not known the null distribution varies with the null hypothesis and even when the parameters are known the power of the procedures is relatively low compared to the regression test.

A major development of EDF tests, which greatly increased their usefulness was due to Stephens (1974, 1976) and Lilliefors (1967). These modifications changed the simple hypothesis tests into ones valid for composite hypotheses and at the same time increased their power. Estimators of the parameters could now be used. However, this meant that percentiles of the test statistic are now dependent on the null distribution.

Stephens (1976) used the following approach to develop composite hypothesis procedures for the normal and exponential (one parameter) distribution. The parameters for the normal are estimated by \bar{y} and s and the parameter for the exponential by $1/\bar{y}$. The observations are ordered such that $y_1 < y_2 < \dots < y_n$. The y_i 's are standardized by the transformation

$$w_i = (y_i - \bar{y})/s$$

for the normal case, and

$$w_i = y_i/\bar{y}$$

for the exponential case. Next the probability integral transformation is made, i.e.,

$$z_i = F(w_i)$$

which for the exponential distribution is simply

$$z_i = 1 - e^{-w_i}.$$

The z_i 's are used for computing the test statistics. Percentiles were given by Stephens (1979) for the following procedures.

$$\text{Cramer-von Mises Test: } W^2 = \sum_{i=1}^n [z_i - (2i-1)/2n]^2 + 1/12n.$$

$$\text{Watson's Test: } U^2 = W^2 - n(\bar{z} - 1/2)^2.$$

Anderson-Darling Test: $A^2 = -\frac{1}{n} \sum_{i=1}^n \{ (2i-1) \ln[z_i(1-z_{n-i+1})] \} - n.$

The test statistics for the normal distribution are computed as follows

$$W^{2*} = W^2(1 + 0.5/n), \quad U^{2*} = U^2(1 + 0.5/n),$$

$$A^{2*} = (1 + 0.75/n + 2.25/n^2)$$

and for the exponential distribution

$$W^{2*} = W^2(1 + 0.16/n), \quad U^{2*} = U^2(1 + 0.16/n),$$

$$A^{2*} = A^2(1 + 0.3/n).$$

Stephens (1974) gave the upper tail percentiles (critical region) for these procedures and Stephens (1976) expanded the available percentiles and included some in the lower tail. Pettitt and Stephens (1976) adapted the EDF procedures for censored samples. They gave percentiles for W^2 , A^2 and U^2 when censoring is in upper, lower or both tails for the normal distribution with parameters unknown.

Stephens (1974) showed by Monte Carlo study that the power of the Anderson-Darling procedure was about as good as the W procedure and recommended it because there were no special constants required to use the test and the percentiles of A^{2*} are just a simple function of sample size and hence no extensive table is needed.

The Kolmogorov-Smirnov procedure is defined as

$$D = \sqrt{n} \max_{1 \leq r \leq n} \left| \frac{r}{n} - F(y_r) \right|$$

where r is the rank of the observation,

$$y_1 < y_2 < \dots < y_n, \text{ and}$$

$F(y_r)$ is the transformed variate.

Exact distributional results for the Kolmogorov-Smirnov tests for the parameter estimated case were obtained by Durbin (1975) for the exponential distribution. Percentiles for sample sizes from 2 to 100 were given.

The use of the Kolmogorov-Smirnov procedure is not recommended since in almost all studies it was shown to have poorer power than either the regression tests or the Anderson-Darling.

Stephens (1977) used the same approach as in his above cited papers to determine the percentiles of W^2 , A^2 and U^2 for testing for the extreme value distribution and Stephens (1979) covered the logistic distribution

$$F(x) = [1 + \exp\{-(x - \epsilon)/\theta\}]^{-1}.$$

A multivariate normal version of W^2 was developed by Malkovich and Afifi (1973). Pettitt (1979) used the W^2 procedure as a test of bivariate normality.

2.3 Specific Property Test. This section describes a number of tests which make use of a specific property of a distribution. These tests are special in that they cannot be generalized to other models. Two of the better known tests for the normal distribution are based on the fact that its standardized third and fourth moments (skewness and kurtosis) are zero and three, respectively. The estimators of these moments have been used as separate tests for normality. These are

$$\sqrt{b_1} = \frac{\sqrt{n} \sum_{i=1}^n (y_i - \bar{y})^3}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)^{1.5}} \quad \text{and} \quad b_2 = \frac{n \sum_{i=1}^n (y_i - \bar{y})^4}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)^2}$$

where \bar{y} is the sample mean. Recently approximations to the distribution of each of these statistics were obtained by D'Agostino and Pearson (1973) and D'Agostino and Tietjen (1973).

These authors used a Johnson distribution to approximate the distribution of the statistics (Johnson distributions can be expressed as transformations of standard normal variables). D'Agostino and Pearson (1973) suggested, in an attempt to improve the power of these procedures, that the two statistics be combined into one omnibus test. They suggested using the statistic

$$K^2 = X^2(\sqrt{b_1}) + X^2(b_2)$$

where $X^2(\cdot)$ was the square of the transformed Johnson variable. Thus, K^2 is the sum of squares of two standard normal variables; however, since $\sqrt{b_1}$ and b_2 are not independent it is not

distributed as chi-square with two degrees of freedom. Bowman and Shenton (1975) provided a series of contours of the 90, 95 and 99 percentiles of K^2 .

Another property of the normal distribution that has been used in test construction is that its entropy exceeds that of any other distribution with the same variance. Vasicek (1976) suggested the statistic

$$K_{m,n} = \frac{n}{2\hat{\sigma}} \left\{ \prod_{i=1}^n (y_{i+m} - y_{i-m}) \right\}^{1/n}$$

where
$$\hat{\sigma}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / n,$$

m is a positive integer less than $n/2$,

$$y_{i-m} = y_1 \quad \text{if } i-m \leq 1,$$

$$y_{i+m} = y_n \quad \text{if } i+m \geq n, \quad \text{and } y_1 < y_2 < \dots < y_n.$$

The author used Monte Carlo procedures to obtain the percentiles of $K_{m,n}$ for $n \leq 50$ and $m = 1(1)5$. Vasicek (1976) did a

Monte Carlo study of the power of the procedure for sample size 20 and $m=3$ and showed that the power of $K_{3,20}$ was similar to that of both A^{2*} and W . He also suggested for optimum power that $m=2$ be used for $n=10$, $m=3$ for $n=20$ and $m=4$ for $n=50$.

Another special property procedure developed as a composite test for the gamma distribution was suggested by Locke (1976).

Using the fact that if x_1 and x_2 are independent, non-degenerate random variables then $x_1 + x_2$ and x_1/x_2 are inde-

pendent if and only if they are gamma variates with the same scale parameter, Locke suggested forming $n/2$ bivariate pairs $[U, V]$

where $U_i = X_{2i-1} + x_{2i}$ and $V_i = \max[x_{2i-1}/x_{2i}, x_{2i}/x_{2i-1}]$. He then uses Kendall's Rank test to see if \tilde{U} and \tilde{V} are independent.

This is one of the few procedures available for testing the composite hypothesis for the gamma distribution.

A variety of test statistics have been developed for the exponential distribution. Defining a gap as the difference between two successive order statistics y_i and y_{i-1} , it is

well known that for an exponential distribution

$$2\lambda S_i = 2\lambda(n-i+1)(y_i - y_{i-1}), \quad i = 1, 2, \dots, n,$$

where $y_0 = \mu = \text{minimum value of } y$, $y_1 < y_2 < \dots < y_n$,

are independent $\chi^2(2)$ variates. Gnedenko *et al.* (1969) suggested using as a two-tailed test for (one-parameter) exponentiality and later investigated by Fercho and Ringer (1972)

$$\Psi(r) = (n-r) \sum_{i=1}^r S_i / r \left(\sum_{i=r+1}^n S_i \right),$$

where $S_1 = n y_1$ and r is a chosen integer. If μ is the origin then $S_1 = n(y_1 - \mu)$.

Since $2\lambda S_i$'s are independent $\chi^2(2)$ then $\Psi(r)$ has an $F(2r, 2(n-r))$ distribution. A modification of this procedure was suggested by Harris (1976) which compared the gaps in the tails to the middle. He recommended

$$\Psi'(r) = (n-2r) \left[\sum_{i=1}^r S_i + \sum_{i=n-r+1}^n S_i \right] / 2r \sum_{i=r+1}^{n-r} S_i.$$

Another version of this procedure was suggested by Lin and Mudholkar (1980). They suggested keeping each tail comparison as one element of a bivariate statistic, i.e., treating

$$F_1 = (n-2r) \sum_{i=1}^r S_i / r \sum_{i=r+1}^{n-r} S_i \quad \text{and} \quad F_2 = (n-2r) \sum_{i=n-r+1}^n S_i / r \sum_{i=r+1}^{n-r} S_i$$

as a bivariate F distribution.

For a composite two-parameter test for exponentiality S_1 can be completely deleted in the above suggested test statistics. The affected terms would lose two degrees of freedom.

The power differentials of the step and leap procedures are small. Neither the Lin-Mudholkar procedure nor the $\Psi'(r)$ procedure appear to improve the Gnedenko procedure. Lin and Mudholkar recommend for optimum power, using $r = [n/10]$ in their procedure while $r = [n/2]$ is generally used with $\Psi(r)$ and $r = [n/4]$ with $\Psi'(r)$ where $[\cdot]$ is the greatest integer symbol. These tests can be used with censored samples since S_i depends only on the two adjacent readings and is an independent chi-square two-degree of freedom variable.

Mann *et al.* (1973) used a similar rationale to develop a test for the extreme value and Weibull distributions. They used the fact that

$$2\ell_i = 2(y_{i+1} - y_i)/E(y_{i+1} - y_i), \quad i = 1, 2, \dots, n-1$$

is asymptotically independent $\chi^2(2)$. Thus, the test statistic for a sample of size n censored at the m th of n observations

$$L(r, s, m, n) = (1/r) \sum_{i=m-r}^{m-1} \ell_i / (1/s) \sum_{i=1}^s \ell_i, \quad r+s+1 \leq m \leq n$$

has asymptotically an $F(2r, 2s)$ distribution. They then suggest an improved form

$$S(r, s, m, n) = (r/s)L / (1 + (r/s)L).$$

They gave percentiles for S and showed that the power of S was better than that for other available tests.

3. CENSORED SAMPLE TEST

In developing a procedure to be used with censored samples the regression of the ordered observations on the expected value of the order statistics will be used. It is possible, with a censored sample to obtain two estimates of σ^2 . One can be based on the slope of the regression line, as is done in the W test procedure, and the other on the residual sum of squares about the regression line. However, in this case we no longer have an analysis of variance rationale since both statistics are only proper estimators of σ^2 under the null hypothesis. The W^* procedure for censored samples compares these two estimates in order to assess whether the hypothesized model is appropriate. When the null hypothesis is false the residual sum of squares will be inflated (due to the poor linear fit) while the slope will not change in a systematic way; it will depend on the alternative model. Thus, the power of the W procedure should be better than W^* and hence the W test should always be used for complete samples.

The null distribution of W^* will depend on the null hypothesis, the actual number of observations and perhaps the sample size. Sections 4 and 5 present some results for the exponential and normal distributions. In what follows it is assumed that the $n-r$ smallest observations in a sample of size n have been recorded.

Following the rationale of Shapiro and Wilk (1965) let $f(x)$ be a standardized density function (location parameter zero and scale parameter one). Let $\tilde{X}' = (x_1, x_2, \dots, x_n)$ be the vector of order statistics of a sample of size n , and

$$\tilde{M}' = E(\tilde{X}') = (m_1, m_2, \dots, m_n)$$

be its expected value. Also let

$$\tilde{V} = (v_{ij}) = E[(\tilde{X}-\tilde{M})(\tilde{X}-\tilde{M})'] \quad i, j = 1, 2, \dots, n$$

be the corresponding covariance matrix of the order statistics. If a linear transformation

$$y_i = \sigma x_i + \mu \quad i = 1, 2, \dots, n$$

is made with $\sigma > 0$ and $-\infty < \mu < \infty$ then the order is preserved and $y_1 < y_2 < \dots < y_n$ are the corresponding transformed order statistics. Also

$$E(y_i) = \sigma m_i + \mu \quad i = 1, 2, \dots, n$$

$$\text{and} \quad \text{cov}(y_i, y_j) = \sigma^2 v_{ij} \quad i, j = 1, 2, \dots, n.$$

Writing $\tilde{Y}' = (y_1, y_2, \dots, y_n)$ and $\tilde{1}' = (1, 1, \dots, 1)$ it follows that

$$E(\tilde{Y}) = \mu \tilde{1} + \sigma \tilde{M}, \quad E[(\tilde{Y} - \mu \tilde{1} - \sigma \tilde{M})(\tilde{Y} - \mu \tilde{1} - \sigma \tilde{M})'] = \sigma^2 \tilde{V}.$$

It is well known that for a positive definite, symmetric matrix such as \tilde{V} , it is always possible to find a non-singular $n \times n$ matrix $\tilde{T} = (t_{ij})$ such that $\tilde{T} \tilde{V} \tilde{T}' = \tilde{I}$. Let $\tilde{Z} = \tilde{T} \tilde{Y}$ where $\tilde{Z}' = (z_1, z_2, \dots, z_n)$. Then

$$E(\tilde{Z}) = \tilde{T}(\mu \tilde{1} + \sigma \tilde{M}) = \mu \tilde{C} + \sigma \tilde{K}$$

$$\text{where } \tilde{C} = \tilde{T} \tilde{1} \text{ and } \tilde{K} = \tilde{T} \tilde{M} \text{ and } \text{cov}(\tilde{Z}) = \sigma^2 \tilde{I}.$$

The matrix \tilde{T} can be chosen so as to be lower triangular and \tilde{T} will be unique if we require the diagonal elements to be positive. Thus,

$$z_1 = t_{11}y_1, \quad z_2 = t_{21}y_1 + t_{22}y_2, \text{ etc.}$$

Because z_i 's are uncorrelated and have a common variance σ^2 , a simple least squares regression of the elements in \tilde{Z} on the elements in \tilde{K} and \tilde{C} is forced through the origin so that the first and second partial regression coefficients provide estimates of σ and μ respectively, and the residual sum of squares provides an estimate of σ^2 . Denoting these as b , a , and s^2 respectively we have for a censored sample of $n-r$ observations the following estimates:

$$b = \frac{\sum_{i=1}^{n-r} z_i k_i \sum_{i=1}^{n-r} c_i^2 - \sum_{i=1}^{n-r} c_i k_i \sum_{i=1}^{n-r} z_i c_i}{d} \quad (3)$$

$$a = \frac{\sum_{i=1}^{n-r} z_i c_i \sum_{i=1}^{n-r} k_i^2 - \sum_{i=1}^{n-r} z_i k_i \sum_{i=1}^{n-r} c_i k_i}{d} \quad (4)$$

$$d = \sum_{i=1}^{n-r} c_i^2 \sum_{i=1}^{n-r} k_i^2 - \left[\sum_{i=1}^{n-r} c_i k_i \right]^2$$

$$s^2 = \sum_{i=1}^{n-r} z_i^2 - b \sum_{i=1}^{n-r} z_i k_i - a \sum_{i=1}^{n-r} z_i c_i. \quad (5)$$

The W^* test is defined by

$$W^* = b^2/s^2. \quad (6)$$

The statistic (6) is origin and scale invariant, hence, it can be used as a composite test for $f(y; \mu, \sigma)$ families. The maximum value of W^* is infinity which occurs when $y_i = m_i$ since $s^2 = 0$ and $b = 1$. Thus, when the observed values equal the expected values the W^* statistic is infinite. Thus, it would be expected that W^* will be a lower tail test.

The minimum value of W^* can be obtained as follows. Since the statistic is scale and origin invariant we can determine the minimum value by letting $a = 0$, $b = 1$ and finding the values of \tilde{Z} that maximize s^2 subject to the above constraints. The constraints $a = 0$, $b = 1$ imply that

$$\sum_{i=1}^{n-r} z_i k_i = \sum_{i=1}^{n-r} k_i^2 \quad \text{and} \quad \sum_{i=1}^{n-r} z_i c_i = \sum_{i=1}^{n-r} c_i k_i. \quad (7)$$

Thus, we want to maximize

$$s^2 = \sum z_i^2 - \sum k_i^2$$

subject to the constraints (7). Since the k_i 's are constants we need only maximize $\sum z_i^2$. This is a convex function which assumes its maximum at one of the vertices of the $(n-2)$ dimensional hypersphere which satisfies the constraints. Therefore, by testing each vertex the maximum value of $\sum z_i^2$ can be found and the minimum point of W^* be computed.

4. APPLICATION TO THE EXPONENTIAL DISTRIBUTION

Let $f(x) = e^{-x}$; $x > 0$. The elements of \tilde{M} and \tilde{Y} take on a simple form for the exponential distribution. Using the results from Kendall and Stuart (1961) and the relationships given in Shapiro and Wilk (1972) the quantities b , a , s^2 in (3), (4), and (5) reduce to

$$b = \frac{\sum_{i=1}^{n-r} y_i + r y_{n-r} - n y_1}{(n-r-1)} \quad (8)$$

$$a = \frac{(n-r)y_1 - \frac{1}{n} \left[\sum_{i=1}^{n-r} y_i + r y_{n-r} \right]}{n-r-1} \quad (9)$$

$$s^2 = \sum_{i=1}^{n-r-1} [2(n-i)(n-i+1)+1] y_i^2 + (r+1)^2 y_{n-r}^2 \quad (10)$$

$$- 2 \sum_{i=2}^{n-r} y_i y_{i-1} (n-i+1)^2 = b \left(\sum_{i=1}^{n-r} y_i + r y_{n-r} \right) - a n^2 y_1.$$

The statistic is defined as $W^* = b^2/s^2$. The minimum value can now be found. We wish to maximize $\sum z_i^2$ subject to the constraints of (7). Now since

$$\sum z_i k_i = \sum z_i, \quad \sum k_i^2 = n-r, \quad \sum z_i c_i = n z_1, \quad \sum c_i k_i = n,$$

the constraints of (7) can be written as

$$\sum z_i = n-r \quad \text{and} \quad z_1 = 1.$$

The vertices in terms of \tilde{z} that satisfy these constraints are

$$(1, n-r-1, 0, \dots, 0)$$

$$(1, 0, n-r-1, 0, \dots, 0)$$

$$\vdots$$

$$(1, 0, \dots, n-r-1).$$

Thus, $s^2 = \sum z_i^2 - \sum k_i^2 = 1 + (n-r-1)^2 - (n-r)$ and the minimum value of W^* is

$$W_{\min}^* = [(n-r-1)(n-r-2)]^{-1}.$$

The distribution of W^* depends only on $n-r$ and not on n . This can be demonstrated when $n-r=3$.

Theorem. The distribution of W^* for $n-r=3$ is given by

$$g(w) = (2w)^{-3/2}; \quad w > 1/2.$$

Proof. Using equations (8), (9) and (10) with $n-r=3$ observations

$$b = [(1-n)y_1 + y_2 + (n-2)y_3]/2$$

$$a = [(3n-1)y_1 - y_2 - (n-2)y_3]/2n$$

and

$$s^2 = [(n-1)y_1 - (2n-3)y_2 + (n-2)y_3]^2/2$$

where the y_i are the $n-r$ smallest observations from a sample of size n from a standard exponential distribution. Make the transformation

$$u_1 = y_1 + y_2 + (n-2)y_3$$

$$u_2 = \frac{(1-n)}{2} y_1 + \frac{y_2}{2} + \frac{(n-2)}{2} y_3$$

$$u_3 = (n-1)y_1 - (2n-3)y_2 + (n-2)y_3.$$

$$f(y_1 y_2 \cdots y_n) = K e^{-\sum_{i=1}^n y_i} \quad 0 < y_1 < y_2 < \cdots < y_n < \infty$$

$$g(y_1 y_2 y_3) = K_1 e^{-\{y_1 + y_2 + (n-2)y_3\}} \quad 0 < y_1 < y_2 < y_3 < \infty$$

$$h(u_1 u_2 u_3) = K_2 e^{-u_1} \quad 0 < |u_3| < 2u_2 < u_1 < \infty$$

$$k(u_2 u_3) = K_3 e^{-2u_2} \quad 0 < |u_3| < 2u_2 < \infty.$$

Let $W = \frac{2u_2^2}{u_3^2} = \frac{b^2}{s^2}$. Integrating out the extraneous variable and

and evaluating the constant yields

$$g(w) = (2w)^{-3/2}; \quad w > 1/2.$$

Thus, we conclude that the distribution for $n-r=3$ is independent of n and only depends on the number of values. We conjecture at this time that this property is true for other values of $n-r$; but we have not been able to derive the exact distribution for samples where $n-r > 3$.

5. APPLICATION TO THE NORMAL DISTRIBUTION

This section uses the censored test rationale for the normal distribution. The matrix \tilde{T} needed for the transformation

$\tilde{Z} = \tilde{T} \tilde{Y}$, has been studied by Shapiro and Wilk (1962) where they showed that \tilde{T} , a lower triangular matrix, is approximately double diagonal. Thus, the z_i , $i = 2, 3, \dots, n$, are represented by the weighted difference of two of the ordered observations. The z_i 's are approximately uncorrelated normal random variables. The test statistic is

$$W^* = b^2/s^2 \quad (11)$$

and equations (3), (4) and (5) are used to calculate the required quantities. The distribution for $n=3$ can be obtained as follows. Using equation (3) it can be shown that

$$b = \frac{t_{33}^2 m_3 (y_3 - y_1) + m_3 (y_1 - y_3) t_{33} t_{31}}{2m_3^2 t_{33} (t_{33} - t_{31})}. \quad (12)$$

Simplification yields

$$b = (y_3 - y_1)/2m_3.$$

Likewise it is easy to show that

$$s^2 = K(y_1 - 2y_2 + y_3)^2, \quad (13)$$

where
$$K = \frac{t_{33}[(1 + 2kt_{33})(t_{33} + t_{31}) - 2k^2 t_{33}]}{2(1 + 2kt_{33})} = .96323$$

$$k = t_{31} + t_{32} + t_{33},$$

using the fact that

$$m_1 = -m_3, \quad m_2 = 0$$

$$t_{11}^2 + t_{21}^2 + t_{31}^2 = t_{33}^2$$

$$t_{31} t_{32} + t_{21} t_{22} - t_{33} t_{32} = 0$$

$$2t_{33} t_{32} + t_{22}^2 + t_{32}^2 = 1.$$

Now applying the transformations

$$\sqrt{3} u_1 = y_1 + y_2 + y_3$$

$$\sqrt{2} u_2 = y_3 - y_1$$

$$\sqrt{6} u_3 = y_3 - 2y_2 + y_1$$

and then changing to polar coordinates and integrating out r yields

$$f(\theta) = 3/\pi, \quad \frac{\pi}{3} < \theta < 2\pi/3.$$

Letting
$$W = \frac{u_2^2}{u_3^2} = \tan^2 \theta \quad \text{then}$$

$$g_1(W) = \frac{3}{2\pi} (W+1)^{-1} W^{-1/2}, \quad 3 \leq W < \infty.$$

To obtain the distribution of W^* we simply note the relationship between b^2 and u_2^2 and s^2 and u_3^2 and obtain

$$g_2(W^*) = \frac{12.41739}{\pi} (W^*+1) W^{*-1/2}; \quad .362 < W^* < \infty.$$

To check the minimum bound we know that we must minimize $\sum z_1^2$ subject to the constraints of the problem and this occurs at a vertex subject to the constraints. An easy way of checking is to use the constraints $b = 1$ and $y_1 \leq y_2 \leq y_3$. Taking the point $(0, z_2, z_3)$ using (12) and noting $z_1 = t_{11} y_1$ and hence $y_1 = 0$ and $y_3 = 2m_3$. Then using (13)

$$s^2 = K(0 - 2y_2 + 2m_3)^2 = 4K(m_3 - y_2)^2.$$

This takes a maximum subject to $0 \leq y_2 \leq 2m_3$ when $y_2 = 0$.

Thus the maximum value of s^2 is $s_{\max}^2 = 4K m_3^2 = 2.76$ and from

(11) $W_{\min}^* = 0.362$ which agrees with the prior result.

The distributional properties of W^* are quite difficult to obtain and it is no longer possible to demonstrate that the percentiles of W^* don't depend on both n and r even for $n=3$. However, examination of the Monte Carlo percentiles which are unpublished indicate that this might be the case. For example, for $n-r=10$ the 5% points for $n=10(1)20$ are .068, .065, .066, .068, .065, .071, .065, .061, .064, .065, .061, and for $n-r=5$ the corresponding 5% points for $n=5(1)10$ are .14, .16, .14, .15, .14, .15.

Further work will be done to determine other properties and obtain an accurate table of percentiles.

6. CONCLUDING REMARKS

This paper has covered some of the more recent tests for distributional assumptions. The coverage is not complete, for example the oldest procedure, chi-square goodness of fit test has been omitted. The authors believe that the procedures cited are among the most powerful available and for the most part can readily be calculated. The users choice of a particular procedure is a personal matter and should depend on the availability of a computer program in order to do the calculations or the

understanding of the individual as to the particular theory behind the test. For most of the competitive procedures presented the power differentials are small. The authors would strongly urge all users of such tests that they augment these analytical procedures with a probability plot. The plot can better describe the data than a single test statistic and the two together provide a good method for analyzing data.

REFERENCES

- Blom, G. (1958). *Statistical Estimates and Transformed Beta Variables*. Wiley, New York.
- Bowman, K. O. and Shenton, L. R. (1975). Omnibus Test Contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, 62, 243-250.
- Brain, C. W. and Shapiro, S. S. (1980). W Test for Weibull distribution. Unpublished manuscript.
- D'Agostino, R. B. (1971). Linear estimation of the Weibull parameters. *Technometrics*, 13, 171-182.
- D'Agostino, R. B. and Pearson, E. S. (1973). Tests for departure from normality: empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, 60, 613-622.
- D'Agostino, R. B. and Tietjen, G. L. (1973). Approaches to the null distribution of $\sqrt{b_1}$. *Biometrika*, 60, 169-173.
- Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika*, 48, 41-55.
- Durbin, J. (1975). Kolmogorov-Smirnov tests when parameters are estimated with application to tests of exponentiality and tests on spacings. *Biometrika*, 62, 5-22.
- Fercho, W. W. and Ringer, L. J. (1972). Small sample power of some tests of constant failure rate. *Technometrics*, 14, 713-724.
- Filliben, J. R. (1975). The probability plot correlation coefficient test for normality. *Technometrics*, 17, 111-117.
- Gnedenko, B. V., Belyayev, Y. K. and Solov'yev, A. D. (1969). *Mathematical Methods of Reliability Theory*. Academic Press, New York.
- Hahn, G. J. and Shapiro, S. S. (1967). *Statistical Models in Engineering*. Wiley, New York.
- Harris, C. M. (1976). A note on testing for exponentiality. *Naval Research Logistic Quarterly*, 23, 169-175.
- Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics Vol. 2*. Hafner, New York.
- LaBrecque, J. (1977). Goodness of fit tests based on non-linearity in probability plots. *Technometrics*, 19, 293-306.
- Lin, C. C. and Mudholkar, G. S. (1980). A test of exponentiality based on the bivariate F distributions. *Technometrics*, 22, 79-82.

- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov tests for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399-402.
- Locke, C. (1976). A test for the composite hypothesis that a population has a gamma distribution. *Communications in Statistics*, 4, 351-364.
- Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association*, 68, 176-179.
- Mann, N. R., Scheuer, E. M. and Fertig, K. W. (1973). A new goodness of fit test for the Weibull distribution or extreme value distribution with unknown parameters. *Communications in Statistics*, 2, 383-400.
- Pearson, K. (1900). On the theory of contingency and its relation to association and normal correlation. *Philosophy Magazine*, 50, 157-175.
- Pettitt, A. N. (1979). Testing for bivariate normality using the empirical distribution function. *Communications in Statistics*, 8, 699-712.
- Pettitt, A. N. and Stephens, M. A. (1976). Modified Gramér-von Mises statistics for censored data. *Biometrika*, 63, 291-298.
- Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.
- Sarkadi, K. (1975). The consistency of the Shapiro-Francia test. *Biometrika*, 62, 445-450.
- Shapiro, S. S. (1964). *An analysis of variance test for normality*. Ph.D. thesis, Rutgers University.
- Shapiro, S. S. (1980). *How to Test Normality and Other Distributional Assumptions*. ASQC Basic References in Quality Control: Statistical Techniques Volume 3. American Society for Quality Control, Milwaukee, Wisconsin.
- Shapiro, S. S. and Francia, R. S. (1972). Approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215-216.
- Shapiro, S. S. and Wilk, M. B. (1962). Approximate uncorrelating transformations for normal order statistics. *Annals of Mathematical Statistics*, 33, 1499.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.
- Shapiro, S. S. and Wilk, M. B. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics*, 14, 355-370.
- Shapiro, S. S. and Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63, 1343-1372.
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69, 730-737.

- Stephens, M. A. (1976). Asymptotic results for goodness of fit statistics with unknown parameters. *Annals of Statistics*, 4, 357-369.
- Stephens, M. A. (1977). Goodness of fit for the extreme value distribution. *Biometrika*, 64, 583-588.
- Stephens, M. A. (1978). On the W test for exponentiality with origin known. *Technometrics*, 20, 33-35.
- Stephens, M. A. (1979). Tests of fit for the logistic distribution. *Biometrika*, 66, 591-595.
- Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B*, 38, 54-59.
- Weisberg, S. (1974). An empirical comparison of the percentage points of W and W'. *Biometrika*, 61, 644-646.

[Received June 1980. Revised October 1980]

A GOODNESS-OF-FIT PROCEDURE FOR TESTING WHETHER A RELIABILITY GROWTH MODEL FITS DATA THAT SHOW IMPROVEMENT

ALAN J. GROSS

Medical University of South Carolina
Charleston, South Carolina 29403 USA

SAMUEL S. SHAPIRO

Florida International University
Miami, Florida 33199 USA

SUMMARY. Reliability growth models that have been discussed in the literature generally assume that an item (such as a piece of electronic equipment) is tested in stages a given number of times, n_i (say, at the i th stage, $i=1,2,\dots,k$). If, at the i th stage, x_i successes are recorded (and hence $n_i - x_i$ failures), the problem is then simply to estimate p_i , the probability of success at the i th stage, $i=1,2,\dots,k$. It is assumed that stages are independent of each other and of time. In this article a general parametric reliability growth model is proposed for p_i and a goodness-of-fit test based on the likelihood ratio criterion is proposed. This test is carried out on data that show reliability growth in stages to determine the adequacy of the fit of the proposed reliability growth model.

KEY WORDS. likelihood ratio test, maximum likelihood estimators, stochastic expansions, chi-square statistics.

1. INTRODUCTION

Consider an item (such as a surface-to-air missile) that undergoes testing in independent stages, where at the i th stage of testing there are n_i items placed on test x_i are successes and $n_i - x_i$ failures. Thus, at the i th stage (assuming Bernoulli trials within each stage) the probability that there are exactly x_i successes and $n_i - x_i$ failures is the binomial model; i.e.,

$$\text{pr}\{X=x_i\} = \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i}, \quad (1)$$

$x_i = 0, 1, \dots, n_i$, wherein p_i is the theoretical probability that any item under test at the i th stage is a success and $1-p_i$ is the corresponding failure probability; $i = 1, 2, \dots, k$.

The object in testing the item in stages is to allow further development and design improvement based on test results which will increase the probability of success from stage to stage. The ultimate goal is to design a reliable item.

If it is assumed that the probability of success increases from stage to stage, subject to random fluctuations, it is important to propose and develop models to assess (estimate) the item reliability at and after each stage of testing and to predict the item reliability of future stages.

Lloyd and Lipow (1962), Gross and Kamins (1968), and Gross and Clark (1975) describe two parametric reliability growth models: the hyperbolic and exponential growth models. These models are described, briefly. The hyperbolic growth model defines p_i as

$$p_i = p_\infty - \alpha/i, \quad (2)$$

where p_∞ , $0 < p_\infty \leq 1$, is the ultimate achievable success probability, and $\alpha > 0$ is a second parameter that quantifies growth between stages 1 and k . The exponential growth model characterizes p_i as

$$p_i = 1 - \alpha_1 \exp(-\alpha_2 i) \quad (3)$$

with $0 \leq \alpha_1 < 1$, and $\alpha_2 > 0$ as the parameters of the model, $i = 1, 2, \dots, k$. The parameter α_2 quantifies growth whereas $1 - \alpha_1$ is the minimum hypothesized success probability for this model.

The parameters of the models (2) and (3) are estimable through the method of maximum likelihood (m.l.). A detailed discussion of the estimation procedure, complete with numerical examples, is found in Gross and Clark (1975; ch. 5). Methods are also developed for predicting the lower confidence limit of p_{k+1} , the success probability at state $k+1$, given k stages of testing.

2. REVIEW OF RELIABILITY GROWTH MODEL LITERATURE

There is an ever-growing bibliography in the area of reliability growth. No attempt is made to give a complete listing of all published work in this field; however, some of the more important articles are cited below.

Lloyd and Lipow (1962) provide one of the earliest discussions of reliability growth models in which they include the hyperbolic and exponential models, (2) and (3), respectively. Other parametric growth models are considered by Berndt (1966), Bresenham (1964), and Zellner and Lee (1965). Gross and Kamins (1968) and Gross and Clark (1975) develop methodology for estimating the hyperbolic and exponential reliability growth model parameters. Their method as previously stated employs m.l. estimation procedures. Furthermore, Gross and Clark demonstrate (in Chapter 5) how the logistic model may be viewed as a reliability growth model and proceed to apply the logistic model in a reliability growth setting.

Nonparametric procedures have also been considered in developing reliability growth models. A key paper in this area was written by Barlow and Scheuer (1966). Their model assumes an item that fails, fails according to an assignable cause that can be corrected while the item undergoes testing; or the item fails due to an inherent cause that cannot be corrected without major redesign of the item. They then obtain the (nonparametric) m.l. estimators of the state-to-stage success probabilities p_i , $i = 1, 2, \dots, k$, under the constraint that p_i 's forms a non-decreasing sequence. The basic problem in obtaining constrained m.l. estimators is first treated by Ayer *et al.* (1955).

Bayes methods have also been used in the development and analyses of reliability growth models. Cozzolino (1966) uses Bayes procedures in making minimum cost decisions on testing times and burn-in procedures for a general class of reliability growth models. If the parameters of a reliability growth model are assumed to be random variables having appropriate (prior) density functions, Pollock (1968) based on this assumption develops a model to project item reliability at some time after the start of testing. The model then predicts reliability of the item after the failure data are observed. Dahiya and Gross (1974) extend an earlier model, Gross (1971), by developing an empirical Bayes procedure to assess reliability growth of an item tested in independent stages. Finally, in the spirit of Bayes, Weinrich and Gross (1978) use a Bayes procedure that utilizes a Dirichlet prior density function on the two failure probabilities and the success probability in the Barlow-Scheuer model to improve upon the estimates of reliability at each stage that were obtained by Barlow and Scheuer (1966).

3. GENERAL PARAMETRIC GROWTH MODELS AND TEST PROCEDURES

Let $p_i(\alpha'_r)$ be the theoretical probability of success of an item at the i th stage of testing where $\alpha'_r = (\alpha_1, \alpha_2, \dots, \alpha_r)$ is a vector of parameters and $i = 1, 2, \dots, k$; $k > r$. The exponential model is an example of the case $r = 2$,

$$p_i(\alpha_1, \alpha_2) = 1 - \alpha_1 \exp(-\alpha_2 i) \quad (3)$$

$i = 1, 2, \dots, k$, and the parameters α_1 and α_2 are the parameters of the model. The unrestricted probability of item success at the i th stage is p_i (independent of α'_r).

A test of the null hypothesis $H_0: p_i = p_i(\alpha'_r)$ against the alternative $H_a: p_i \neq p_i(\alpha'_r)$, p_i unrestricted, is carried out by the likelihood ratio criterion. Thus if

$$\Lambda = \prod_i \frac{p_i^{x_i}(\hat{\alpha}'_r) q_i^{n_i - x_i}(\hat{\alpha}'_r)}{\hat{p}_i^{x_i} \hat{q}_i^{n_i - x_i}}, \quad (4)$$

where $\hat{\alpha}'_r$ is the m.l. estimator vector of α'_r , $\hat{p}_i = x_i/n_i$, the m.l. estimator of p_i , $q_i(\hat{\alpha}'_r) = 1 - p_i(\hat{\alpha}'_r)$, and $\hat{q}_i = 1 - \hat{p}_i$,

$i = 1, 2, \dots, k$. The statistic $-2 \ln \Lambda$ is approximately distributed as a chi-square variable with $(k-r)$ degrees of freedom when H_0 is true. Thus, H_0 is rejected at level α only if

$$-2 \ln \Lambda > \chi_{k-r, 1-\alpha}^2,$$

where $\chi_{k-r, 1-\alpha}^2$ is the $(1-\alpha)100$ percentage point of the chi-square distribution with $(k-r)$ degrees of freedom.

In order to simplify the test procedure, $-2 \ln \Lambda$ is now studied more closely.

$$\begin{aligned} -2 \ln \Lambda = & 2 \sum_i x_i \ln x_i / \{n_i p_i (\hat{q}'_r)\} \\ & + 2 \sum_i (n_i - x_i) \ln(n_i - x_i) / \{n_i q_i (\hat{q}'_r)\}. \end{aligned} \quad (5)$$

Let $d_i = x_i - n_i p_i (\hat{q}'_r)$, $i = 1, 2, \dots, k$. Then, it follows that $-d_i = (n_i - x_i) - n_i q_i (\hat{q}'_r)$. Thus, $-2 \ln \Lambda = 2 \sum_i (n_i - x_i) \ln\{1 + d_i / (n_i - x_i)\}^{-1} + 2 \sum_i x_i \{\ln(1 - d_i / x_i)\}^{-1}$.

Theorem 1. Statistic W_1 , where

$$W_1 \equiv \sum_i d_i^2 [\{n_i p_i (\hat{q}'_r)\}^{-1} + \{n_i q_i (\hat{q}'_r)\}^{-1}], \quad (6)$$

is asymptotically distributed as a chi-square variables with $k-r$ degrees of freedom, when H_0 is true.

Proof. Consider the stochastic expansions of $\ln\{1 + d_i / (n_i - x_i)\}^{-1}$ and $\ln(1 - d_i / x_i)^{-1}$. It follows that,

$$\begin{aligned} -2 \ln \Lambda = & 2 \sum_i x_i \{A + A^2/2 + A^3/3 + \dots\} \\ & + 2 \sum_i (n_i - x_i) \{B + B^2/2 + B^3/3 + \dots\} \end{aligned}$$

where $A = d_i / x_i$ and $B = d_i / (n_i - x_i)$. Thus, one finds

$$-2 \ln \Lambda = \sum_i d_i^2 \left\{ x_i^{-1} + (n_i - x_i)^{-1} \right\} \\ + \text{terms in } \sum_i d_i^a \left\{ x_i^{-(a-1)} + (n_i - x_i)^{-(a-1)} \right\}$$

$a \geq 3$. Now, for each a ,

$$\sum_i d_i^2 \left\{ x_i^{-(a-1)} + (n_i - x_i)^{-(a-1)} \right\} \\ = \sum_i \left[(d_i^2/x_i) (d_i/x_i)^{a-2} + d_i^2/(n_i - x_i) \cdot \{d_i/(n_i - x_i)\}^{a-2} \right].$$

Consider the term d_i/x_i which can be written as

$$d_i/x_i = \frac{\{x_i/n_i - p_i(\hat{\alpha}'_r)\}}{x_i/n_i} - \frac{\{p_i(\hat{\alpha}'_r) - p_i(\alpha'_r)\}}{x_i/n_i}.$$

When H_0 is true, $x_i/n_i \xrightarrow{P} p_i(\alpha'_r)$ and $p_i(\hat{\alpha}'_r) \xrightarrow{P} p_i(\alpha'_r)$. Thus, $d_i/x_i \xrightarrow{P} 0$. Similarly, $d_i/(n_i - x_i) \xrightarrow{P} 0$. Furthermore,

$$d_i^2/x_i = d_i^2/\{n_i p_i(\hat{\alpha}'_r) + d_i\} = d_i^2/[n_i \{p_i(\hat{\alpha}'_r) + d_i/n_i\}].$$

Since $d_i/n_i \xrightarrow{P} 0$, for n_i large $d_i^2/x_i \xrightarrow{P} d_i/n_i p_i(\hat{\alpha}'_r)$ and similarly, $d_i^2/(n_i - x_i) \xrightarrow{P} d_i^2/n_i q_i(\hat{\alpha}'_r)$. Both $d_i^2/n_i p_i(\hat{\alpha}'_r)$ and $d_i^2/n_i q_i(\hat{\alpha}'_r)$ are bounded with probability 1. Thus,

$$\sum_i d_i^a \{x_i^{-(a-1)} + (n_i - x_i)^{-(a-1)}\} \xrightarrow{P} 0 \quad \text{as } n_i \rightarrow \infty.$$

Let $W_1 \equiv -2 \ln \Lambda$. Thus, as $n_i \rightarrow \infty$, $i=1,2,\dots,k$, the distribution of W_1 is chi-square with $k-r$ degrees of freedom when H_0 is true. Q.E.D.

Note that W_1 is analogous to the statistic for goodness-of-fit that Berkson (1953) introduces. Note that no rationale is given for the Berkson statistic. Such a rationale is the *raison d'être* for this article.

A generalization of Theorem 1 is obtainable. Suppose the null hypothesis under test is $H_0: p_{i0} \equiv p_i(\alpha'_s, \alpha'_{r-s})$ vs. $H_a: p_i \equiv p_i(\alpha'_r)$. That is, the test is now that r -s of the parameters of the model $p_i(\alpha'_r)$ are specified under H_0 as opposed to the alternative that no parameters are specified, $i=1,2,\dots,k; k \geq r \geq s$.

Theorem 2. Define $\hat{p}_{i0} \equiv p_i(\hat{\alpha}'_s, \hat{\alpha}'_{r-s})$ and $\hat{p}_i \equiv p_i(\hat{\alpha}'_r)$; $i=1,2,\dots,k$. The statistic W_2 is asymptotically distributed as a chi-square variable with r -s degrees of freedom when H_0 is true; where

$$W_2 \equiv \sum_i \left[\frac{x_i(\hat{p}_i - \hat{p}_{i0})(3\hat{p}_i - \hat{p}_{i0})}{\hat{p}_{i0}^2} + \frac{(n_i - x_i)(\hat{q}_{i0} - \hat{q}_i)(\hat{q}_{i0} - 3\hat{q}_i)}{\hat{q}_{i0}^2} \right]$$

$$\hat{q}_i = 1 - \hat{p}_i, \hat{q}_{i0} = 1 - \hat{p}_{i0}; i=1,\dots,k.$$

Proof. Similar to that of Theorem 1.

4. AN EXAMPLE APPLIED TO THE HYPERBOLIC RELIABILITY GROWTH MODEL

The hyperbolic reliability growth model, given by (2) was fitted to the following clinical trial situation that it discussed by Gross and Clark (1975, pp. 173-176). Patients with acute leukemia were tested in five stages, 10 new patients selected at each stage. Between stages, medication was modified so that the proportion of patients achieving remission from stage to stage should increase. The proportions of patients achieving remission at each state are 3/10, 6/10, 7/10, and 8/10, for stages 1 through 5. The hyperbolic reliability growth model, (2), was proposed in this case. Gross and Clark show that the m.l. estimates of p_∞ and α are, respectively, $\hat{p}_\infty = 0.8891$ and $\hat{\alpha} = 0.5892$. This leads to the following table of numbers of observed and expected successes at each stage.

TABLE 1: Observed and expected number of successes at each of five stages of the leukemia clinical trial.

Stage	Observed Successes	Expected Successes (Hyperbolic Model)
1	3	2.999
2	6	5.845
3	7	6.927
4	7	7.418
5	8	7.713

When H_0 is true the statistic W_1 is approximately a chi-square variable with three degrees of freedom. Now

$$\begin{aligned}
 W_1 = & (3-2.999)^2[(2.999)^{-1} + (7.001)^{-1}] \\
 & + (6-5.845)^2[(5.845)^{-1} + (4.055)^{-1}] \\
 & + (7-6.927)^2[(6.927)^{-1} + (3.073)^{-1}] \\
 & + (7-7.418)^2[(7.418)^{-1} + (2.582)^{-1}] \\
 & + (8-7.713)^2[(7.713)^{-1} + (2.287)^{-1}] = 0.140.
 \end{aligned}$$

Now $\chi^2_{3,0.90} = 6.25$, indicating the fit of the hyperbolic model to these data is excellent.

REFERENCES

- Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 26, 641-647.
- Barlow, R. E. and Scheuer, E. M. (1966). Reliability growth during a development testing program. *Technometrics*, 8, 53-60.
- Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *Journal of the American Statistical Association*, 48, 565-599.
- Berndt, G. D. (1966). *Estimating a Monotonically Changing Binomial Parameter*. Office of Operation Analysis, Strategic Air Command, Offutt Air Force Base, Nebraska.

- Bresenham, J. E. (1964). *Reliability Growth Models*. Technical Report No. 74, Department of Statistics, Stanford University.
- Cozzolino, J. M., Jr. (1966). *The Optimal Burn-in Testing of Repairable Equipment*. Technical Report No. 23, MIT Operations Research Center, Cambridge, Massachusetts.
- Dahyia, R. C. and Gross, A. J. (1974). Adaptive exponential smoothing models for reliability estimation. *IEEE Transactions on Reliability*, R-23, 332-334.
- Gross, A. J. (1971). The application of exponential smoothing to reliability assessment. *Technometrics*, 13, 877-883.
- Gross, A. J. and Clark, V. A. (1975). *Survival Distributions; Reliability Applications in the Biomedical Sciences*. Wiley, New York.
- Gross, A. J. and Kamins, M. (1968). Reliability assessment in the presence of reliability growth. *Annals of Assurance Sciences: 1968 Annual Symposium on Reliability*, 406-416.
- Lloyd, D. K. and Lipow, M. (1962). *Reliability: Management, Methods and Mathematics*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Pollock, S. M. (1968). A Bayesian reliability growth model. *IEEE Transactions on Reliability*, R-17, 187-198.
- Weinrich, M. C. and Gross, A. J. (1978). The Barlow-Scheuer reliability growth model from a Bayesian viewpoint. *Technometrics*, 20, 249-254.
- Zellner, A. and Lee, T. H. (1965). Joint estimation of relationships involving discrete random variables. *Econometrica*, 33, 382-394.

[Received May 1980. Revised September 1980]

CHI-SQUARE GOODNESS-OF-FIT TESTS BASED ON DEPENDENT OBSERVATIONS

KAMAL C. CHANDA*

Department of Mathematics
Texas Tech University
Lubbock, Texas 79409 USA

SUMMARY. An attempt has been made in this article to investigate the sampling properties of standard goodness-of-fit chi-square tests based on interdependent observations which are obtained from a strictly stationary and strong-mixing random process. We conclude that the null distributions of the test statistics are largely determined by the multivariate probability structure of the process. A few examples are used to illustrate this point.

KEY WORDS. goodness-of-fit tests, chi-square tests, strictly stationary and strong mixing processes, linear processes.

1. INTRODUCTION

Let $\{X_t: t = 0, \pm 1, \dots\}$ be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.), with a common distribution function (d.f.) G . If one wants to test various kinds of hypotheses H concerning G , based on a sample X_1, \dots, X_n , one would, probably, like to use the classical goodness-of-fit tests based on such statistics as the chi-square statistics, Kolmogorov-Smirnov statistics, or other statistics which exploit the characterization properties of G under H . The simplest and the most commonly used among these are the chi-square goodness-of-fit tests of which a comprehensive account is

* Research was supported in part by the Air Force Office of Scientific Research under Contract No. F-49620-79-C-0194.

available in Watson (1959). Problem arises, however, when it is known *a priori* that the observations X_1, \dots, X_n are not independent but G is still the d.f. of X_t , implying strict stationarity for $\{X_t\}$, in which case it is essential that some information regarding the structure of dependence of the sequence $\{X_t\}$ is available, before an appropriate chi-square test can be formulated. If we use the standard chi-square statistics in such situations, the resulting tests might have properties which are substantially different from those expected, for both small and large samples. Bartlett (1950) and Patankar (1954) have illustrated this point in some detail for Markov chains and stationary Gaussian sequences when H is simple.

Since a meaningful probability treatment of strict stationarity will, possibly, require some assumption under which observations far separated from each other are nearly independent, we shall assume throughout that $\{X_t\}$ is either a strong-mixing process with mixing coefficients $\alpha(v)$, ($v=1,2,\dots$) (for definition of such processes see Rozanov, 1967, p. 180) or a

linear process defined by $X_t = \sum_{v=0}^{\infty} g_v W_{t-v}$ where $\{W_t: t=0, \pm 1, \dots\}$ is a sequence of i.i.d. r.v.'s and the infinite sum converges to X_t in some stochastic sense. There are features of similarity common to these two processes but a linear process is not, in general, strong-mixing, although it can be so under some additional conditions (see Chanda, 1974, and Gorodetskii, 1977). These two processes together constitute possibly the widest such class of strictly stationary processes encountered in the literature.

We further assume that $G(x) = G(x; \underline{\theta})$, where $\underline{\theta}$ is a parameter vector and that G is absolutely continuous with a probability density function (p.d.f.) g . The support of G is the interval (a, b) which is divided into r class intervals $B_j = (a_{j-1}, a_j)$, ($1 \leq j \leq r$, $a_0 = a$, $a_r = b$). We consider two situations. (I) G is fully specified under H , in which case the class probabilities are given by

$$p_j = G(a_j; \underline{\theta}) - G(a_{j-1}; \underline{\theta}), \quad (1 \leq j \leq r). \quad (1)$$

(II) G is fully specified under H except that $\underline{\theta}$ is unknown and is estimated by a statistic $\hat{\underline{\theta}}$ in which case we estimate the class probabilities by

$$\hat{p}_j = G(a_j; \hat{\theta}) - G(a_{j-1}; \hat{\theta}), \quad (1 \leq j \leq r). \quad (2)$$

Sections 2 and 3 contain the main body of results followed by general discussion. The proofs of the results stated in the form of theorems and comments are withheld and will be reported elsewhere.

2. CHI-SQUARE GOODNESS-OF-FIT TESTS

In the classical tradition we define the chi-square statistics by

$$X^2 = \sum_{j=1}^r (n_j - np_j)^2 / np_j \quad \text{in case (I)} \quad (3)$$

$$\hat{X}^2 = \sum_{j=1}^r (n_j - n\hat{p}_j)^2 / n\hat{p}_j \quad \text{in case (II)} \quad (4)$$

where n_j = number of observations X_t in the class interval B_j , $(1 \leq j \leq r)$.

2.1 Case I: p_j $(1 \leq j \leq r)$ are known. The following theorem describes the asymptotic behavior of the chi-square statistic.

Theorem 2.1. Let X^2 be as defined in (3), and let $\{X_t: t = 0, \pm 1, \dots\}$ be a strictly stationary strong-mixing process with $\alpha(v) = O(v^{-1-\epsilon})$ for some $\epsilon > 0$. Assume that $p_j \geq c$, $1 \leq j \leq r$, for some $c > 0$. Let $\Lambda = [\lambda_{ij}]$ where

$$\lambda_{ij} = \sum_{v=-\infty}^{\infty} \lambda_{ij}(v),$$

$$\lambda_{ij}(v) = [p_{ij}(v) - p_i p_j] (p_i p_j)^{-1/2}, \quad (5)$$

$$p_{ij}(v) = P(X_1 \in B_i, X_{1+v} \in B_j), \quad v=0, \pm 1, \dots,$$

$$p_{ij}(0) = \begin{cases} p_i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then, as $n \rightarrow \infty$,

$$L(X^2) \rightarrow L\left(\sum_{j=1}^{r-1} \tau_j \chi_j^2\right), \quad (6)$$

where χ_j^2 are mutually independent r.v.'s each with a chi-square distribution with 1 degree of freedom and $\tau_1, \dots, \tau_{r-1}, 0$ are the eigenvalues of Λ .

Theorem 2.1'. Let $\{X_t\}$ be a strictly stationary linear process defined by $X_t = \sum_{v=0}^{\infty} g_v W_{t-v}$ where $\{W_t\}$ is a sequence of i.i.d. r.v.'s with $E(|W_1|^\delta) < \infty$ for some $\delta > 0$ and $\sum_{v=0}^{\infty} v |g_v|^\lambda < \infty$, $\lambda = \delta/2(1 + \delta)$. Let the remaining conditions of Theorem 2.1 be satisfied. Then (6) holds.

Observe that although $\lambda_{ij}(0)$ ($1 \leq i, j \leq r$) are known because p_j 's are given, the quantities $\lambda_{ij}(v)$ ($v \neq 0, 1 \leq i, j \leq r$) are not usually known. If these are given as additional conditions under the hypothesis H then we replace X^2 by

$$X^{*2} = Z' \tilde{\Lambda}^- Z, \quad (7)$$

where $\tilde{\Lambda}^-$ is the (unique) g-inverse of $\tilde{\Lambda}$ (for definition of g-inverse see Graybill, 1976, p. 24). It can be shown that

$$L(X^{*2}) \rightarrow L(\chi^2(d)), \quad (8)$$

where $\chi^2(d)$ has a chi-square distribution with d degrees of freedom, $d = \text{rank}(\tilde{\Lambda}) \leq r-1$.

Note that if $\{X_t\}$ is a sequence of i.i.d. r.v.'s then $\Lambda = \bar{I} - p p'$ where $p' = [p_1^{1/2} \dots p_r^{1/2}]$, so that $\tau_j = 1$, $1 \leq j \leq r-1$, and $d = r-1$. It follows immediately that $\Lambda^- = \Lambda$ and since $p'Z = 0$, we have that $X^{*2} = Z'\Lambda^-Z = Z'\Lambda Z = Z'Z = X^2$.

In general, it is difficult to compute λ_{ij} even if analytic expressions for $p_{ij}(v)$ are known. For example, if $\{X_t\}$ is Gaussian with zero mean, unit variance and autocorrelation function $\rho(v)$ then it is well known that for $v \neq 0$,

$$p_{ij}(v) = \sum_{u=0}^{\infty} \rho^u(v) c_{iu} c_{ju} / u! \text{ where } c_{iu} \text{ (} 1 \leq i \leq r, 0 \leq u < \infty \text{)}$$

are given in Section 4. Even if $\rho(v)$ is completely known λ_{ij} is hard to compute. It is, therefore, desirable to find out if there exist estimates $\tilde{\lambda}_{ij}$ (of λ_{ij}) which can be used to modify X^2 so that this new statistic has a simple asymptotic null distribution. The result of the following theorem shows that such a task is possible to accomplish.

Theorem 2.2. Define

$$\tilde{\lambda}_{ij}(v) = (n-v)^{-1} \sum_{t=1}^{n-v} Y_i(t) Y_j(t+v) \quad (9)$$

$$\tilde{\lambda}_{ij}(-v) = \tilde{\lambda}_{ji}(v),$$

$$Y_i(t) = [I_i(X_t) - p_i] p_i^{-1/2}, \quad (1 \leq i \leq r, 1 \leq t \leq n),$$

$$I_i(x) = \begin{cases} 1 & \text{if } x \in B_i \\ 0 & \text{otherwise,} \end{cases}$$

$0 \leq v \leq n-1$, $1 \leq i, j \leq r$. Let $\{\ell_n\}$ be a sequence of positive integers such that $\ell_n \rightarrow \infty$ but $\ell_n^3/n \rightarrow 0$ as $n \rightarrow \infty$ and let the mixing coefficient $\alpha(v)$ satisfy the condition of Theorem 2.1. Define

$$\tilde{\Lambda} = [\tilde{\lambda}_{ij}], \quad (10)$$

$$\tilde{\lambda}_{ij} = \sum_{|v| \leq \ell_n} \tilde{\lambda}_{ij}(v)$$

$$\tilde{X}^{*2} = \tilde{Z}' \tilde{\Lambda} \tilde{Z}$$

where $\tilde{\Lambda}^-$ is the g -inverse of $\tilde{\Lambda}$. Then as $n \rightarrow \infty$

$$L(\tilde{X}^{*2}) \rightarrow L(\chi^2(d)), \quad (11)$$

where $d = \text{rank } (\Lambda)$.

Theorem 2.2'. Let $\{X_t\}$ be a linear process satisfying the conditions of Theorem 2.1'. Then if the remaining conditions of Theorem 2.2 are satisfied (11) holds.

2.2 Case II: p_j ($1 \leq j \leq r$) are estimated. We now consider the situation where p_j are unknown. We assume that θ is a parameter vector with s elements ($s < r$). Let $\hat{\theta}$ be estimated by the estimator $\hat{\theta}$. We assume that $\hat{\theta} = n^{1/2}(\hat{\theta} - \theta) = n^{-1/2} \sum_{t=1}^n H_t + o_p(1)$ where $H_t = h(X_t)$ ($h' = [h_1, \dots, h_s]$) for some function h_j ($1 \leq j \leq s$), $E(H_t) = 0$ and $E(|H_t|^{2+n}) < \infty$ for some $\eta > 0$. Let $Y'_t = [Y_1(t), \dots, Y_r(t)]$. Write $p_{ij} = (\partial p_i / \partial \theta_j) p_j^{-1/2}$, $P = [p_{ij}]$, $E(H_t Y'_{t+v}) = \tilde{A}(v) = [a_{ij}(v)]$, $E(H_t, H'_{t+v}) = \tilde{R}(v) = [\rho_{ij}(v)]$. We then have the following result.

Theorem 2.3. Let the conditions of Theorem 2.1 hold true and let $|p_{ij}| < M$ for every i, j and θ , M being a finite positive constant independent of θ . Assume that $\eta > 4/\epsilon$. Then

$\tilde{A} = \sum_{v=-\infty}^{\infty} \tilde{A}(v)$, and $\tilde{R} = \sum_{v=-\infty}^{\infty} \tilde{R}(v)$ converge absolutely and, as $n \rightarrow \infty$,

$$L(\hat{\tilde{X}}^2) \rightarrow L \left[\sum_{j=1}^{r-1} \tau_j^* \chi_j^2 \right], \quad (12)$$

where τ_j^* ($1 \leq j \leq r-1$) and 0 are the eigenvalues of $\tilde{\Lambda}^* = \tilde{\Lambda} - \tilde{P}\tilde{A} - \tilde{A}'\tilde{P}' + \tilde{P}\tilde{R}\tilde{P}'$ and χ_j^2 ($1 \leq j \leq r-1$) are mutually independent r.v.'s each with a chi-square distribution with one degree of freedom.

Note that if $\{X_t\}$ is a sequence of i.i.d. r.v.'s then

$$\tilde{\Lambda}^* = \tilde{\Lambda}^*(0) = \tilde{I} - \tilde{p}\tilde{p}' - \tilde{P}\tilde{A}(0) - \tilde{A}'(0)\tilde{P}' + \tilde{P}\tilde{R}(0)\tilde{P}'. \quad (13)$$

A reformulation of Theorem 2.3 in terms of linear processes is too complicated and hence withheld.

3. EFFECTS OF ESTIMATION OF PARAMETERS ON $\tilde{\Lambda}^*$

We now investigate different methods of estimating $\tilde{\theta}$ and the effect of such estimation on the matrix $\tilde{\Lambda}^*$.

A. Suppose $\hat{\tilde{\theta}}$ is the 'restricted' maximum likelihood estimator of $\tilde{\theta}$ derived as a suitable solution of

$$\sum_{t=1}^n \partial \ln g(X_t; \tilde{\theta}) / \partial \theta_j = 0, \quad 1 \leq j \leq s. \quad (14)$$

Then under a set of regularity conditions (see Crámer, 1946, pp. 500-501) on g it can be established that strict stationarity and strong mixing property of $\{X_t\}$ will imply that

$$\tilde{\delta} = n^{-1/2} \sum_{t=1}^n J^{-1}_{\tilde{\theta}} L(X_t) + o_p(1), \quad (15)$$

where

$$\tilde{J} = [\tilde{j}_{uv}]$$

$$\tilde{j}_{uv} = -E(\partial^2 \ln g(X_t; \theta) / \partial \theta_u \partial \theta_v), \quad 1 \leq u, v \leq s,$$

$$\tilde{L}' = [L_1 \dots L_s],$$

$$L_u(x) = \partial \ln g(x; \tilde{\theta}) / \partial \theta_u.$$

Thus for this particular case we can write $\tilde{h}(x) = \tilde{J}^{-1} \tilde{L}(x)$.

B. Suppose $\hat{\tilde{\theta}}$ is the 'restricted' modified minimum chi-square estimator of $\tilde{\theta}$ defined as an appropriate solution of

$$\sum_{i=1}^r n_i \partial \ln p_i / \partial \theta_j = 0 \quad 1 \leq j \leq s. \quad (16)$$

Then the following result holds.

Theorem 3.1. Let $\{X_t\}$ satisfy the condition of Theorem 2.1 or 2.1'. Let $\tilde{\theta}$ belong to an admissible set Θ and let its true value θ^0 be an interior point of Θ . We assume that (i) for every $\tilde{\theta}, \theta^* \in \Theta, \tilde{\theta} \neq \theta^*, p_i(\tilde{\theta}) \neq p_i(\theta^*)$ for at least one i ($1 \leq i \leq r$), (ii) $\partial p_i(\theta) / \partial \theta_j$ ($1 \leq j \leq s, 1 \leq i \leq r$) exist and are continuous at $\tilde{\theta}^0$ and (iii) $\tilde{Q} = \tilde{P}' \tilde{P}$ is nonsingular at $\tilde{\theta} = \tilde{\theta}^0$. Then there exists a consistent root $\hat{\tilde{\theta}}$ of (16) such that

$$n^{1/2}(\tilde{\theta} - \tilde{\theta}^0 - \tilde{Q}_0^{-1} \tilde{M}_n^0) \xrightarrow{P} 0 \quad (17)$$

where $\tilde{Q}_0 = \tilde{Q}(\tilde{\theta}^0)$, $\tilde{M}_n^0 = [\tilde{M}_{1n}^0 \dots \tilde{M}_{sn}^0]$, $\tilde{M}_{jn} = \sum_{i=1}^r (n_i/n) (\partial \ln p_i / \partial \theta_j)$ and $\tilde{M}_{jn}^0 = \tilde{M}_{jn}(\tilde{\theta}^0)$.

It follows, immediately, that for this case $\tilde{H}_t = (\tilde{P}' \tilde{P})^{-1} \tilde{P}' \tilde{Y}_t$ and therefore

$$\hat{\Lambda}^* = (\hat{I} - \hat{S})\hat{\Lambda}(\hat{I} - \hat{S}), \quad (18)$$

where \hat{I} is an identity matrix of order r and $\hat{S} = P(P'P)^{-1}P'$.

C. Let $g(x; \theta) = f((x - \theta_1)/\theta_2)$ and assume that f is symmetric about zero. Estimate θ_1 by the sample median $\hat{\theta}_1$ and θ_2 by $\hat{\theta}_2 = \text{interquartile range}/2c$ where $\int_{-\infty}^c f(x)dx = 3/4$.

Then it can be proved that for linear processes defined earlier (see Chanda, 1976; and Dutta and Sen, 1971)

$$n^{1/2}(\hat{\theta}_1 - \theta_1) = (n^{1/2}f(0))^{-1} \sum_{t=1}^n \phi_1(X_t) + o_p(1), \quad (19)$$

$$n^{1/2}(\hat{\theta}_2 - \theta_2) = (2n^{1/2}cf(c))^{-1} \sum_{t=1}^n \phi_2(X_t) + o_p(1),$$

where $\phi_1(x) = 1/2 - \alpha(x - \theta_1)$, $\phi_2(x) = \alpha(x - \xi_1) - \alpha(x - \xi_3) + 1/2$, $\alpha(x) = 1$ if $x \leq 0$ and $\alpha(x) = 0$ if $x > 0$ and $\xi_1 = \theta_1 - c\theta_2$, $\xi_3 = \theta_1 + c\theta_2$. Therefore, in this case

$$H'_t = [(f(0))^{-1}\phi_1(X_t), (2cf(c))^{-1}\phi_2(X_t)]. \quad (20)$$

It is not difficult to prove that the result of Theorem 2.3 applies here.

4. A FEW EXAMPLES

In this section we derive expressions for $\hat{\Lambda}$ and $\hat{\Lambda}^*$ for a few special cases.

I. Assume that $\{X_t\}$ is Gaussian with $E(X_t) = \mu$, $V(X_t) = 1$, $\text{Cov}(X_t, X_{t+v}) = \rho(v) = O(v^{-1-\epsilon})$ for some $\epsilon > 0$, and

$1 + 2 \sum_{v=1}^{\infty} \cos v \lambda \rho(v) > 0$ for all λ , $(-\pi \leq \lambda \leq \pi)$. The last

two conditions guarantee strong mixing property of $\{X_t\}$ with

mixing coefficient $\alpha(v) = O(v^{-1-\varepsilon})$ (see Theorem 10.1 and 10.2 in Rozanov, 1967).

(a) Let $\hat{\mu} = \bar{X} = n^{-1} \sum_{t=1}^n X_t$. This leads to $H_t = X_t - \mu$ and

$$\tilde{\Lambda}^* = \tilde{\Lambda} - \gamma \tilde{P} \tilde{P}', \quad (21)$$

where $\gamma = \sum_{v=-\infty}^{\infty} \rho(v)$, $\tilde{\Lambda} = [\lambda_{ij}]$, $\lambda_{ij} = \lambda_{ij}(0) + 2(p_i p_j)^{-1/2}$
 $\sum_{u=1}^{\infty} d_u c_{iu} c_{ju} / u!$, $\lambda_{ij}(0) = -(p_i p_j)^{1/2}$ if $i \neq j$, $\lambda_{ii}(0) = 1 - p_i$,
 $d_u = \sum_{v=1}^{\infty} \rho^u(v)$, $c_{iu} = \phi^{(u-1)}(\alpha_i) - \phi^{(u-1)}(\alpha_{i-1})$, $\phi(x) = (2\pi)^{-1/2}$
 $\exp(-x^2/2)$, $\alpha_i = a_i - \mu$, $p_i = \phi(\alpha_i) - \phi(\alpha_{i-1}) = c_{i0}$, $\phi(x) =$
 $\int_{-\infty}^x \phi(y) dy$, $\tilde{P}' = [k_1 \dots k_r]$, $k_i = c_{i1} p_i^{-1/2}$. If $\rho(v) = 0$ for all
 $v \neq 0$ then $\gamma = 1$ and $\tilde{\Lambda} = [\lambda_{ij}(0)]$. This result was first
 derived by Chernoff and Lehmann (1954). If $\{X_t\}$ is Gaussian-
 Markov with $\rho(v) = \rho^v$, $|\rho| < 1$ ($v \geq 0$) then $\gamma = (1 + \rho)/(1 - \rho)$
 and $d_u = \rho^u / (1 - \rho^u)$.

(b) Let $\hat{\theta} = \text{median of the sample } X_1, \dots, X_n$. Then $H_t =$
 $(2\pi)^{1/2} (1/2 - \alpha(X_t - \mu))$ (see (19)) and

$$\tilde{\Lambda}^* = \tilde{\Lambda} - \tilde{P} \tilde{A} - \tilde{A}' \tilde{P}; + \tilde{P} \tilde{R} \tilde{P}', \quad (22)$$

where $\tilde{\Lambda}$, \tilde{P} are as in (21), $\tilde{A}' = [a_{11} \dots a_{1r}]$ where $a_{1i} =$
 $a_{1i}(0) - p_i^{-1/2} \sum_{m=0}^{\infty} d_{2m+1} h_m c_{i,2m+1} / (2m+1)!$, $h_m = (-1)^m (2m)! / m! 2^m$,
 and

$$a_{1i}(0) = \begin{cases} (\pi p_1/2)^{1/2} & , \text{ if } \alpha_{i-1} \geq 0, \\ -(\pi/2p_1)^{1/2}(1 - p_1 - 2\phi(\alpha_{i-1})), & \text{ if } \alpha_{i-1} < 0 < \alpha_i, \\ -(\pi p_1/2)^{1/2} & , \text{ if } \alpha_i \leq 0. \end{cases}$$

And, finally.

$$R = \pi/2 + 2 \sum_{v=1}^{\infty} \arcsin \rho(v).$$

II. Let $\{X_t\}$ be Gaussian with $E(X_t) = \mu$, $V(X_t) = \sigma^2$ and covariance function same as in I above. Let $\hat{\mu} = \bar{X}$ and

$$\hat{\sigma} = [n^{-1} \sum_{t=1}^n (X_t - \bar{X})^2]^{1/2}. \text{ Then } \underline{H}'_t = [X_t - \mu, \{(X_t - \mu)^2 - \sigma^2\}/2\sigma] \text{ and}$$

$$\underline{\Lambda}^* = \underline{\Lambda} - \underline{P}\underline{D}\underline{P}' \quad (23)$$

where $\underline{\Lambda}$ is the same as in (21), $\underline{P} = [p_{ij}]$, $p_{i1} = -(\phi(\alpha_i) - \phi(\alpha_{i-1}))/\sigma$, $p_{i2} = -(\alpha_i\phi(\alpha_i) - \alpha_{i-1}\phi(\alpha_{i-1}))/\sigma$, $\alpha_i = (a_i - \mu)/\sigma$, $(1 \leq i \leq r)$, $\underline{D} = \sigma^2/2 \begin{pmatrix} 2\gamma & 0 \\ 0 & \delta \end{pmatrix}$, γ is as defined in (21) and

$$\delta = \sum_{v=-\infty}^{\infty} \rho^2(v). \text{ If } \mu \text{ is known and we need to estimate only}$$

σ then

$$\underline{\Lambda}^* = \underline{\Lambda} - \underline{\gamma}\underline{P}\underline{P}' \quad (24)$$

where $\underline{\gamma} = \sigma^2\delta/2$ and $\underline{P}' = [p_{11} \dots p_{r1}]$, with $p_{i1} = -[\alpha_i\phi(\alpha_i) - \alpha_{i-1}\phi(\alpha_{i-1})]/\sigma$, $\alpha_i = (a_i - \mu)/\sigma$.

III. Let $\{X_t\}$ be a Cauchy-Markov process defined by

$$X_t - \mu = \rho(X_{t-1} - \mu) + W_t, \quad (25)$$

$-\infty < \mu < \infty$, $|\rho| < 1$, where $\{W_t\}$ is a sequence of i.i.d. r.v.'s with d.f. F , and $F(w) = 1/2 + \arctan w/\pi$, $(-\infty < w < \infty)$. Note that $\{X_t\}$ is a linear process satisfying the conditions of Theorem 2.1'. It is not difficult to establish that for $v > 0$, $|p_{ij}(v) - p_{ij}| \leq Mv\rho_0^v$ where $\rho_0 = |\rho|$ and M is a finite positive constant not depending on v or ρ . It is also easy to see that for $v > 0$

$$\lambda_{ij}(v) = (p_i p_j)^{-1/2} \int_{\alpha_{i-1}}^{\alpha_i} f(x) [F(w_j(x;0)) - F(w_j(x;\rho))] dx$$

where $w_j(x;\rho) = [(1 - \rho_0^v)^2 + (\alpha_j - \rho^v x)(\alpha_{j-1} - \rho^v x)]/(\alpha_j - \alpha_{j-1})$,

$f(x) = F'(x)$, $\lambda_{ij}(-v) = \lambda_{ji}(v)$, and $\alpha_i = (a_i - \mu)/\sigma$. Also for

$v > 0$ and $1 \leq i \leq r$, $a_{1i}(v) = -\pi p_i^{-1/2} \int_{-\infty}^0 f(x) [F(w_i(x;0) -$

$F(w_i(x;\rho))] dx$, $a_{1i}(-v) = p_i^{-1/2} \int_{\alpha_{i-1}}^{\alpha_i} f(x) \arctan(\rho^v x/\gamma_v) dx$,

$$\gamma_v = 1 - \rho_0^v,$$

$$a_{1i}(0) = \begin{cases} \pi p_i^{1/2}/2 & , \text{ if } \alpha_{i-1} \geq 0, \\ -\pi p_i^{1/2}/2[1 - p_i - 2F(\alpha_{i-1})], & \text{ if } \alpha_{i-1} < 0 < \alpha_i, \\ -\pi p_i^{1/2}/2 & , \text{ if } \alpha_i \leq 0. \end{cases}$$

Also for all $v > 0$, $R(v) = \pi \int_0^{\infty} f(x) \arctan(\rho^v x/\gamma_v) dx = R(-v)$, $R(0) = \pi^2/4$.

TABLE 1. Values of τ_j, τ_j^* ($1 \leq j \leq r$) for Gaussian-Markov process for selected values of ρ and r , ($p_1 = p_2 = \dots = p_r = r^{-1}$).

ρ	τ_1	τ_1^*	τ_2	τ_2^*	τ_3	τ_3^*	τ_4	τ_4^*	τ_5	τ_5^*
$r = 2$										
-0.9	0.22	0.19								
-0.8	0.32	0.25								
-0.6	0.48	0.32								
-0.4	0.62	0.35								
-0.2	0.79	0.36								
0.0	1.00	0.36								
0.2	1.32	0.37								
0.4	1.86	0.38								
0.6	2.98	0.43								
0.8	6.39	0.68								
0.9	13.28	1.18								
$r = 4$										
-0.8	2.71	2.71	0.87	0.87	0.18	0.08				
-0.6	1.49	1.49	0.97	0.97	0.33	0.11				
-0.4	1.15	1.15	0.99	0.99	0.50	0.13				
-0.2	1.03	1.03	1.00	1.00	0.71	0.14				
0.0	1.00	1.00	1.00	1.00	1.00	0.14				
0.2	1.43	1.03	1.03	1.00	1.00	0.14				
0.4	2.16	1.15	1.15	1.01	1.01	0.15				
0.6	3.62	1.49	1.49	1.04	1.04	0.18				
0.8	8.04	2.71	2.71	1.26	1.24	0.27				
$r = 6$										
-0.8	3.30	3.30	1.10	1.10	1.00	1.00	0.72	0.72	0.15	0.05
-0.6	1.68	1.68	1.01	1.01	1.00	1.00	0.92	0.92	0.29	0.06
-0.4	1.22	1.22	1.00	1.00	1.00	1.00	0.98	0.98	0.47	0.07
-0.2	1.05	1.05	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.08
0.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08
0.2	1.46	1.05	1.05	1.00	1.00	1.00	1.00	1.00	1.00	0.08
0.4	2.23	1.22	1.22	1.02	1.02	1.00	1.00	1.00	1.00	0.09
0.6	3.78	1.68	1.68	1.12	1.10	1.01	1.01	1.00	1.00	0.10
0.8	8.45	3.30	3.30	1.62	1.59	1.10	1.10	1.01	1.01	0.15

5. NUMERICAL RESULTS

This section contains some numerical details which will illustrate the effect of the structural dependence of $\{X_t\}$ on the validity of X^2 - and \hat{X}^2 -tests defined in (3) and (4) respectively. Since there are infinite variations in the choice of r , p_i ($1 \leq i \leq r$) and G , it is difficult to decide which few individual cases we should consider numerically. A simple choice would be the case where G is $N(0,1)$, $p_1 = p_2 = \dots = p_r = r^{-1}$ and $|\rho(v)| \leq M\rho_0^v$ ($v > 0$) where M is a finite positive constant and $0 \leq \rho_0 < 1$. Any $\{X_t\}$ which is stationary Gaussian and is autoregressive of a finite order will be a linear process satisfying the property of $\rho(v)$, and will be a suitable example to consider. Again since the asymptotic validity of the X^2 - and \hat{X}^2 -tests in such cases depend only on d_u (see (21)) we may adequately represent the general Gaussian structure by the simple Gaussian-Markov process with $\rho(v) = \rho^v$ ($v \geq 0$) whereby $d_u = \rho^u / (1 - \rho^u)$. For case I we assume that $\{X_t\}$ is Gaussian-Markov with $E(X_t) = 0$, $V(X_t) = 1$, $p_1 = p_2 = \dots = p_r = r^{-1}$. For the corresponding Case II we assume that $\{X_t - \mu\}$ is Gaussian-Markov with the property as in Case I and we estimate μ by \bar{X} . The numerical values of τ_j, τ_j^* ($1 \leq j \leq r$) ($\tau_r = 0 = \tau_r^*$) for selected values of ρ and r are given in Table 1.

It appears that τ_j^* are most stable and contribute more to the validity of the chi-square test than τ_j ($1 \leq j \leq r$), over a wide range of values of ρ .

REFERENCES

- Bartlett, M. S. (1950). The frequency goodness-of-fit test for probability chains. *Proceedings of the Cambridge Philosophical Society*, 47, 86-95.
- Chanda, K. C. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, 11, 401-408.
- Chanda, K. C. (1976). Some comments on sample quantiles for dependent observations. *Communications in Statistics - Theory and Methods*, A5(14), 1385-1392.
- Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit. *Annals of Mathematical Statistics*, 25, 579-586.

- Crámer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- Dutta, K. and Sen, P. K. (1971). On the Bahadur representation of sample quantiles in some stationary multivariate autoregressive processes. *Journal of Multivariate Analysis*, 1, 186-198.
- Graybill, F. A. (1976). *Theory and Applications of the Linear Model*. Duxbury Press, North Scituate, Massachusetts.
- Gorodetskii, V. V. (1977). On the strong mixing property for linear sequences. *Theory of Probability and its Applications*, 22, 411-413.
- Patankar, V. N. (1954). The goodness-of-fit of frequency distributions obtained from stochastic processes. *Biometrika*, 50, 450-462.
- Rozanov, Yu. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.
- Watson, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics*, 15, 440-468.

[Received June 1980. Revised October 1980]

AN ASYMPTOTICALLY DISTRIBUTION-FREE GOODNESS-OF-FIT TEST FOR FAMILIES OF STATISTICAL DISTRIBUTIONS DEPENDING ON TWO PARAMETERS

FORTUNATO PESARIN *

Institute of Statistics
University of Padova
35100 Padova, Italy

SUMMARY. This paper concerns a general goodness-of-fit test for families of statistical distributions depending on two nuisance parameters and satisfying some general conditions. The null distribution of the test statistic does not depend on the nuisance parameters and is asymptotically distribution-free. The paper gives a short table of its critical values and a table of its power against some alternatives.

KEY WORDS. Asymptotically distribution-free test, goodness-of-fit test, nuisance parameters, families of statistical distributions, weighted Cramér-von Mises test.

1. INTRODUCTION

Consider a family V_F of continuous random variables with cumulative distribution functions $F(x; \theta_1, \theta_2)$, where the functional form of F is fixed and (θ_1, θ_2) is a pair of real nuisance parameters. Often, but not always, θ_1 and θ_2 are location and scale parameters, respectively.

We will present a general goodness-of-fit test for the null hypothesis H_0 that $G(x)$ belongs to V_F , $G(x)$ being the

* This work is a part of research supported by the Italian National Research Council (Grant n. CT 77.0153.10).

c.d.f. of the observable r.v. X . The null distribution of the test statistic (a) is (θ_1, θ_2) -independent for every sample size $n > 2$ and (b) is asymptotically V_F -independent, i.e. the test is asymptotically distribution-free.

2. CONDITIONS AND TEST

The validity of the test assumes the following conditions:

- i) there are two monotone continuous real functions q and g , not depending on (θ_1, θ_2) , and two real functions a and b , depending on (θ_1, θ_2) , such that $q(F(x; \theta_1, \theta_2)) = a + b \cdot g(x)$ for every member of V_F ;
- ii) The relation between (a, b) and (θ_1, θ_2) is one to one and the functions $\theta_1(a, b)$ and $\theta_2(a, b)$ are continuous;
- iii) the derivative $q'(u) = dq(u)/du$ is almost surely uniformly continuous and different from 0, with respect to every member of V_F , in the interval $[0, 1]$.

Let Ω_F denote the set of all r.v.'s having the variances of the transforms $q(F(X; \theta_1, \theta_2))$ and $g(X)$ finite and positive, for all (θ_1, θ_2) . In this context the functions q and g characterize the family V_F and linearize the c.d.f.'s of its members.

Families of distributions satisfying the properties i) to iii) include, for instance: Normal, Lognormal, Logistic, Pareto, Weibull, Exponential, Gumbel, etc.

In the case where (θ_1, θ_2) are location and scale parameters, so that $F(x; \theta_1, \theta_2) = F_0((x - \theta_1)/\theta_2)$, we have

$$q = F_0^{-1}, \quad g(x) = x, \quad a = -\theta_1 \theta_2^{-1}, \quad b = \theta_2^{-1}.$$

Let $X'_1 \leq X'_2 \leq \dots \leq X'_n$ be the order statistics corresponding to a random sample drawn from $G(x)$ and consider the quantities:

$$q_i = q(i/(n+1)), \quad q'_i = q'_i(i/(n+1)), \quad g_i = g(X'_i), \\ w_i = (n+1)(n+2)/[i(n+1-i)(q'_i)^2], \quad i = 1, \dots, n.$$

The test we propose is based on the statistic

$$D^* = \min_{(a,b)} \{ \sum_i (q_i - a - b \cdot g_i)^2 w_i \} = (1-r^2) D_q \quad \text{where}$$

$$D_q = \sum q_i^2 w_i - (\sum q_i w_i)^2 / \sum w_i, \quad D_g = \sum g_i^2 w_i - (\sum g_i w_i)^2 / \sum w_i,$$

$C_{q,g} = \sum q_i g_i w_i - (\sum q_i w_i \cdot \sum g_i w_i) / \sum w_i$, $r^2 = C_{q,g}^2 / (D_q \cdot D_g)$, are respectively the variance of q_i , the variance of g_i , the covariance between q_i and g_i and the squared linear correlation coefficient, all weighted with w_i . The statistic D^* corresponds to the weighted residual variance from the least squares linear regression.

The test statistic we propose is given by

$$D = D^*(n+1)/(n-2).$$

Obviously large values of D are significant.

3. THE TEST'S PROPERTIES

Assume H_0 is true. We first establish the properties mentioned at the end of Section 1.

To prove property (a) we simply note that r^2 is unchanged under nondegenerate linear transforms so that it is (a,b) -independent. Then it is also (θ_1, θ_2) -independent by conditions i) and ii), and so D is.

To prove property (b) we consider the r.v. $q(F(X'_i; \theta_1, \theta_2)) = a + b \cdot g(X'_i) = q(U'_i)$, $i = 1, \dots, n$, where U'_i are, by the continuity of F , the order statistics from the uniform distribution $U(0,1)$. By the uniform continuity of q' , if n is sufficiently large, we have that $q(U'_i) \approx q_1 + (U'_i - i/(n+1)) \cdot q'_1$, $i = 1, \dots, n$. So $q(U'_i)$ is, asymptotically, a linear transform of U'_i and their joint density function is proportional to that of U'_i . Thus the test's components

$$(q_i - a - b \cdot g_i)^2 w_i \approx (U'_i - i/(n+1))^2 (n+1)(n+2)/[i(n+1-i)]$$

are jointly asymptotically not dependent in distribution on q and then not dependent on F . This assures that D is asymptotically V_F -independent in distribution.

TABLE 1: Critical values for some selected families calculated by a Monte Carlo procedure with 2000 replications.

V_F	n :	10	20	40	100	200	α
Exponential		.840	.952	.918	.914	.946	.05
		1.193	1.293	1.342	1.319	1.335	.01
Logistic		.705	.800	.762	.702	.688	.05
		1.055	1.194	1.077	1.006	.952	.01
Normal		.688	.731	.713	.705	.684	.05
		.999	1.101	1.077	.974	.949	.01
Weibull		.687	.785	.754	.706	.687	.05
		1.002	1.108	1.075	1.002	.953	.01

TABLE 2: Explicit expressions for the quantities q_i , w_i , g_i for the families: a) Exponential, b) Logistic, c) Normal, d) Weibull, e) Pareto, f) Gompertz. Φ is the standard normal c.d.f.

V_F	q_i	w_i	g_i
a)	$\ln[(n+1)/(n+1-i)]$	$(n+2)(n+1-i)/[i(n+1)]$	X'_i
b)	$\ln[i/(n+1-i)]$	$(n+2)i(n+1-i)/(n+1)^3$	X'_i
c)	$z_i : \Phi(z_i) = i/(n+1)$	$\frac{(n+1)(n+2)}{2\pi i(n+1)} \exp(-z_i^2)$	X'_i
d)	$\ln(\ln \frac{n+1}{n+1-i})$	$\frac{(n+2)(n+1-i)}{i(n+1)} (\ln \frac{n+1-i}{n+1})^2$	$\ln X'_i$
e)	$\ln[(n+1)/(n+1-i)]$	$(n+2)(n+1-i)/[i(n+1)]$	$\ln X'_i$
f)	$-\ln[\ln((n+1)/i)]$	$\frac{i(n+2)}{(n+1-i)(n+1)} (\ln \frac{1}{n+1})^2$	X'_i

TABLE 3: Power of D with $\alpha = .05$ for the families: a) Exponential, b) Logistic, c) Normal, d) Weibull; against some indicated alternatives. Values calculated by a Monte Carlo procedure with 500 replications. Dash means test not applicable.

Alternative	V_F :	$n = 20$				$n = 40$			
		a)	b)	c)	d)	a)	b)	c)	d)
Exponential		.050	.090	.720	.050	.045	.125	.972	.050
Logistic		.802	.042	.144	-	.990	.054	.190	-
Normal		.712	.028	.046	-	.990	.046	.043	-
Weibull		.050	.090	.720	.050	.045	.125	.972	.050
Lognormal		.248	.206	.880	.176	.368	.396	.992	.358
Chi-square (1)		.460	.180	.962	.092	.774	.216	.996	.136
Half-normal		.080	.022	.292	.120	.148	.034	.620	.136
Half-logistic		.026	.044	.496	.084	.082	.100	.838	.094

D is a consistent test against $H_1 = \{X \text{ in } \Omega_F - V_F\}$. To prove this we observe the analogy between D and the weighted Cramér-von Mises goodness-of-fit test, which is asymptotically independent of the nuisance parameters and is based on the statistic

$$A^2 = n \sum [F(\hat{X}_1^i; \hat{\theta}_1, \hat{\theta}_2) - i/n]^2 / [F_1(1-F_1)],$$

where $(\hat{\theta}_1, \hat{\theta}_2)$ is a "good" estimate of (θ_1, θ_2) . We know that the test based on A^2 is consistent, provided that (θ_1, θ_2) admits a consistent estimator. We have that the continuity of q , the consistency of the least squares estimators in our regularity conditions, the continuity of $\theta_1(a,b)$ and $\theta_2(a,b)$, and the consistency and the weighted Cramér-von Mises test assure the consistency of D .

The set of alternatives not contained in H_1 can be divided into two subsets: the one for which the test is not consistent and the other for which it is not applicable. This last situation occurs for instance in cases where $g(x)$ is not real or is not defined on some set of real numbers having positive probability with respect to some members of V_F .

From Table 1 we observe that the critical values are practically n -invariant and substantially the same, with respect to different families of statistical distributions, for even moderate values of n .

REFERENCES

- Csörgő, S. and Stacho, L. (1980). A step toward an asymptotic expansion for the Cramér-von Mises statistic. In *Colloquia Mathematica Societatis Janos Bolyai, Analytic Function Methods in Probability Theory*, B. Gyres, ed. North-Holland Publishing Company, Amsterdam, Pages 53-65.
- Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.
- Pesarin, F. and Azzalini, A. (1976). Verifica di ipotesi funzionale non parametrica. In *L'Elaborazione Automatica*, 2, 53-77.
- Pesarin, F. (1977). A goodness-of-fit test for families of random variables depending on two parameters with censored data. *Recent Developments in Statistics*, I. R. Barra et al., ed. North Holland. Pages 567-570.
- Pesarin, F. (1978). Controllo statistico di curve di affidabilità dipendenti da due parametri. In *Atti del Decimo*

Convegno Nazionale della Associazione Italiana per il Controllo della Qualità, 77-83.

- Pettitt, A. N. (1978). Generalized Cramér-von Mises statistics for the gamma distribution. *Biometrika*, 65, 232-235.
- Sarkadi, K. (1975). The consistency of the Shapiro-Francia test. *Biometrika*, 62, 445-450.
- Shapiro, S. S. and Francia, R. S. (1972). Approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215-225.
- Stephens, M. A. (1974). EDF Statistic for goodness-of-fit and some comparisons. *Journal of the American Statistical Association*, 69, 730-737.

[Received August 1980. Revised October 1980]

CONDITIONALITY PROPERTIES FOR THE BIVARIATE LOGARITHMIC DISTRIBUTION WITH AN APPLICATION TO GOODNESS OF FIT

A. W. KEMP

School of Mathematical Sciences
University of Bradford
Bradford, BD7 1DP ENGLAND

SUMMARY. The paper gives new modes of genesis for the multivariate and homogeneous bivariate logarithmic distributions. Marginality and conditionality properties are derived (including, e.g., $X|Y > X$, $X|Y - X = n$), and are applied to the problem of grouping frequencies for a chi-square goodness-of-fit test for the homogeneous bivariate logarithmic distribution. The proposed procedure is objective and impartial with respect to rows and columns.

KEY WORDS. Bivariate logarithmic series distribution, multivariate logarithmic series distribution, modes of genesis, chi-squared goodness-of-fit, objective grouping, homogeneous distributions.

1. INTRODUCTION

The multivariate logarithmic series distribution (LSD-m) was introduced into the literature by Khatri (1959). Subsequent papers, Patil and Bildikar (1967), Chatfield, Ehrenberg and Goodhardt (1966), see also Johnson and Kotz (1969), stressed its close relationship to the multivariate negative binomial distribution (NBD-m), just as the famous Fisher, Corbett and Williams (1943) paper emphasized the affinity between the univariate logarithmic series distribution (LSD-1) and the univariate negative binomial distribution (NBD-1).

This paper deals with the bivariate logarithmic series distribution (LSD-2) and the bivariate negative binomial distribution (NBD-2) as bivariate forms of LSD-m and NBD-m. The no-

tation for these and other closely related distributions is given in Section 2, which also explains the concept of homogeneity for a multivariate discrete distribution. Section 3 is devoted to the genesis of LSD-m and LSD-2; we shall show that as well as NBD-m and NBD-2 based models there are many other models not related to the Fisherian-type limiting process. Marginal and conditional distributions for LSD-2 are given in Section 4; these results concern not only $X|Y = y$ and $X|X + Y = n$, but also $X|X > Y$, $X|X < Y$, $X|X = Y$, $X|X - Y = n$, $X|Y - X = n$.

Finally, Section 5 shows how these results can be applied to the problem of carrying out a chi-squared goodness-of-fit test for a bivariate distribution. A grouping procedure is proposed which is entirely objective and moreover is impartial with respect to rows and columns. A numerical illustration is given using data known to be bivariate logarithmic.

2. NOTATION

Consider \underline{x} , \underline{p} , \underline{z} and \underline{d} in R^m . Let $\underline{x} = (x, y, \dots)$ be an m -component random vector where all components have non-negative integer support. Let $\underline{p} = (a, b, \dots)$ be an m -component parameter vector with components c satisfying $0 < c \leq 1$. Let $\underline{z} = (s, t, \dots)$ be an m -component vector of generating variables, and let $\underline{d} = (1, 1, \dots)$.

\underline{x} has the *multinomial* distribution (MD-m) if its p.g.f. is of the form

$$(1 - \underline{p} \cdot \underline{d} + \underline{p} \cdot \underline{z})^n$$

where $\underline{p} \cdot \underline{d} < 1$ and n is a positive integer. The binomial distribution is MD-1.

\underline{x} has the *singular multinomial* distribution (SMD-m) if its p.g.f. is of the form $(\underline{p} \cdot \underline{z})^n$, where $\underline{p} \cdot \underline{d} = 1$ and n is a positive integer.

\underline{x} has the *multivariate negative binomial* distribution, also called the negative multinomial distribution (NBD-m), if its p.g.f. is of the form

$$(1 - \underline{p} \cdot \underline{d})^k / (1 - \underline{p} \cdot \underline{z})^k = (1 + \underline{q} \cdot \underline{d} - \underline{q} \cdot \underline{z})^{-k}$$

where $0 < k$, $\underline{p} \cdot \underline{d} < 1$ and $\underline{q} = \underline{p} / (1 - \underline{p} \cdot \underline{d})$. The univariate negative binomial distribution is NBD-1, whilst the bivariate negative binomial distribution is NBD-2. Note that this is the Bates

and Neyman (1952) form of the distribution, introduced originally by Guldberg (1934), see also Sibuya, Yoshimura and Shimizu (1964); it is not the more general version of Edwards and Gurland (1961).

x has the *multiple Poisson* distribution (MP-m) if its p.g.f. is of the form

$$\exp\{\lambda(\tilde{p} \cdot \tilde{z} - \tilde{p} \cdot \tilde{d})\}, \quad 0 < \lambda.$$

This is just the convolution of m independent Poisson distributions.

x has the *multivariate logarithmic series* distribution (LSD- \tilde{m}) if its p.g.f. is of the form

$$\log(1 - \tilde{p} \cdot \tilde{z}) / \log(1 - \tilde{p} \cdot \tilde{d}) \quad \text{where } \tilde{p} \cdot \tilde{d} < 1.$$

The univariate and bivariate logarithmic distributions are LDS-1 and LSD-2, where LSD-2 is the distribution studied by Patil and Bildikar (1967) and Chatfield *et al.* (1966). It is not the bivariate logarithmic distribution of Patil and Joshi (1968).

x has the *modified multivariate logarithmic* distribution (MLSD- \tilde{m}) if its p.g.f. is of the form

$$1 - \theta + \theta \log(1 - \tilde{p} \cdot \tilde{z}) / \log(1 - \tilde{p} \cdot \tilde{d}), \quad \tilde{p} \cdot \tilde{d} < 1, \quad 0 < \theta < 1.$$

Again the univariate and bivariate modified logarithmic distributions are MLSD-1 and MLSD-2.

The multivariate version of the convolution of a geometric distribution with either a binomial with $m = 1$ or alternatively a pseudo-binomial type I (with exponent parameter equal to unity) will be called the *multivariate binomial convolution* distribution (BCD-m) and will be defined as having the p.g.f.

$$\{\theta + (1 - \theta)\tilde{p} \cdot \tilde{z}\} / \{1 + \phi - \phi\tilde{p} \cdot \tilde{z}\}$$

where $\tilde{p} \cdot \tilde{d} = 1$, $0 < \theta < 1 + \phi$, $0 < \phi$. In the univariate case (BCD-1), the distribution is $(+B) * (-B)$ when $0 < \theta < 1$, $0 < \phi$, and $(-B) * (pB_1)$ when $1 < \theta < 1 + \phi$, see Kemp (1979).

A multivariate discrete distribution will be said to be *homogeneous* if its p.g.f. contains the generating variables only in the form of a linear function within a function of a function. All the distributions defined above are therefore homogeneous. Note that NBD-2 is homogeneous whilst the Gurland-Edwards bivariate negative binomial is not. Also the usual bivariate Poisson with p.g.f.

$$\exp\{a(s-1) + b(t-1) + c(st-1)\}$$

is not homogeneous.

Section 4 onwards will use $G_{x,y}(s,t) = H(as + bt)$ as the p.g.f. of an homogeneous bivariate distribution of X and Y . $G_x(s)$ will denote the marginal distribution of X , and $G_{x|y=n}(s)$ will denote the distribution of X conditional on Y taking the value n .

3. GENESIS OF LSD-m AND LSD-2

Many modes of genesis for LSD-m are extensions of those for LSD-1. Instances of LSD-2 and LSD-m will occur therefore in the many areas (ecology, linguistics, marketing, meteorology, etc.) where LSD-1 has been found appropriate. For lucid reviews of models for LSD-1 see Nelson and David (1967) and Boswell and Patil (1971).

3.1 Fisher's limit model. Analogously to the univariate case, LSD-m is the limit as $k \rightarrow 0$ of the origin-truncated NBD-m, see Patil and Bildikar (1967). Hence, any model which leads to NBD-m can be expected to provide a model for LSD-m; for models for NBD-m see e.g. Bates and Neyman (1952), Sibuya, Yoshimura and Shimizu (1964) and Neyman (1965).

3.2 Fisher's mixture model. The NBD-1 is the gamma-mixed Poisson distribution where the gamma distribution has parameter k . When the parameter $k \rightarrow 0$ there are two possibilities - either the Poisson can be truncated after mixing, or it can be truncated before mixing. The multivariate analogues are as follows:

3.2a Truncation after mixing. Let the mixing distribution be the exponential integral distribution with probability density function (p.d.f.)

$$e^{-w/\alpha} / E_1(\epsilon/\alpha)w, \quad \epsilon \leq w < \infty.$$

The mixed MP-m distribution has then the p.g.f.

$$\int_{\epsilon}^{\infty} \frac{\exp\{w(\tilde{p} \cdot \tilde{z} - \tilde{p} \cdot \tilde{d})\} e^{-w/\alpha}}{E_1(\epsilon/\alpha)w} dw = E_1[\epsilon/\alpha - \epsilon(\tilde{p} \cdot \tilde{z} - \tilde{p} \cdot \tilde{d})] / E_1(\epsilon/\alpha).$$

Truncation of the origin then gives the multivariate distribution with p.g.f.

$$\frac{\gamma + \log(\varepsilon/\alpha + \varepsilon \tilde{p} \cdot \tilde{d} - \varepsilon \tilde{p} \cdot \tilde{z}) + S_1 - \gamma - \log(\varepsilon/\alpha + \varepsilon \tilde{p} \cdot \tilde{d}) - S_2}{\gamma + \log(\varepsilon/\alpha) + S_3 - \gamma - \log(\varepsilon/\alpha + \varepsilon \tilde{p} \cdot \tilde{d}) - S_2} = \frac{\log\{1 - \tilde{p} \cdot \tilde{z}/(1/\alpha + \tilde{p} \cdot \tilde{d})\} + S_1 - S_2}{\log\{1 - \tilde{p} \cdot \tilde{d}/(1/\alpha + \tilde{p} \cdot \tilde{d})\} + S_3 - S_2}$$

where $E_1(\cdot)$ is the exponential integral, γ is Euler's constant, $S_1 = \sum_{n \geq 1} (-\varepsilon/\alpha - \varepsilon \tilde{p} \cdot \tilde{d} + \varepsilon \tilde{p} \cdot \tilde{z})^n / n!$, $S_2 = [S_1]_{\tilde{z}=(0,0,\dots)}$ and $S_3 = [S_1]_{\tilde{z}=\tilde{d}}$. Finally letting $\varepsilon \rightarrow 0$ gives LSD-m. More informal univariate versions of this model have been given by Kendall (1948), McCloskey (1965) and Boswell and Patil (1971).

3.2b *Truncation before mixing.* Suppose now that the mixing distribution has p.d.f.

$${}_1F_1(1; 2; w) e^{-w/\alpha} / {}_2F_1(1, 1; 2; \alpha) = (e^w - 1) e^{-w/\alpha} / (-w \cdot \log(1 - \alpha)) \quad , \quad 0 \leq w < \infty, \quad 0 < \alpha.$$

The mixed origin-truncated MP-m distribution will then have the p.g.f.

$$\int_0^\infty \frac{e^{w \tilde{p} \cdot \tilde{z}} - 1}{e^w - 1} \cdot \frac{(e^w - 1) e^{-w/\alpha}}{(-w) \log(1 - \alpha)} dw = \log(1 - \alpha \tilde{p} \cdot \tilde{z}) / \log(1 - \alpha).$$

The univariate version of this model was first suggested by Kendall (1948) and was given by Boswell and Patil (1971).

3.3 *Kendall's birth-and-death models.* Kendall showed that LSD-1 is a limiting form as $k \rightarrow 0$ of the immigration processes with p.g.f.'s

$$[(\beta - \mu)/(\beta \Lambda - \mu - \beta \Lambda s + \beta s)]^{k/\beta} \quad \text{where } \Lambda = e^{(\beta - \mu)\tau}, \quad \beta \neq \mu,$$

and $(1 + \beta T - \beta T)^{-k/\beta} \quad \text{where } \beta = \mu,$

as in the Fisher limit model.

Kendall also examined the birth-and-death processes (no immigration) having BCD-1 with p.g.f.

$$\frac{\mu \lambda - \mu - \mu \lambda s + \beta s}{\beta \lambda - \mu - \beta \lambda s + \beta s} \quad \text{where } \lambda = e^{(\beta - \mu)\tau}, \quad \beta \neq \mu,$$

and
$$\frac{\beta t + (1 - \beta t)s}{1 + \beta t - \beta ts} \quad \text{where } \beta = \mu;$$

for other processes leading to this distribution, see e.g. Kemp (1979). Kendall showed that if t has a uniform distribution on $(-T, 0)$ then the resultant distribution is MLSD-1 with p.g.f.

$$1 + \frac{1}{\beta T} \log(1 - x) \left[1 - \frac{\log(1 - xs)}{\log(1 - x)} \right],$$

where $x = (\beta\Lambda - \beta)/(\beta\Lambda - \mu)$, $\Lambda = e^{(\beta-\mu)\tau}$, if $\beta \neq \mu$,

and $x = \beta T/(1 + \beta T)$ if $\beta = \mu$.

More generally, if z has the BCD-m with p.g.f.

$$(\mu\lambda - \mu - \mu\lambda p \cdot z + \beta p \cdot z)/(\beta\lambda - \mu - \beta\lambda p \cdot z + \beta p \cdot z)$$

$$\text{or } (\beta t + p \cdot z - \beta t p \cdot z)/(1 + \beta t - \beta t p \cdot z)$$

where $p \cdot d = 1$, $\lambda = e^{(\beta-\mu)\tau}$ and t is uniform on $(-T, 0)$, then the resultant distribution is MLSD-m.

3.4 Kendall-Patil binomial mixture. The MLSD-m has been presented in Section 2 as a binomial ($n = 1$) mixture of LSD-m. As Patil and Bildikar (1967) have pointed out, MLSD-m is also a mixed MD-m where the MD-m parameter n has an LSD-1. This is a multivariate analogue of a result given in Kendall (1948).

3.5 Quenouille's Poisson-log negbin relationship. Quenouille (1949) gave an important result relating the Poisson, logarithmic and negative binomial distributions. He showed that NBD-1 arises as the sum of n independent LSD-1 variables where n has MP-1. Recently Kemp (1978) has shown that the mathematics underlying this relationship are the same as those for the Fisher limiting derivation of LSD-1. The mathematical relationship between NBD-1 and LSD-1 has here two quite distinct interpretations.

The multivariate analogue of Quenouille's relationship is that NBD-m is the sum of n independent LSD-m variables where n has the Poisson distribution, see Philippou and Roussas (1974).

3.6 MLSD-m as a limit distribution. Consider now the NBD-m with p.g.f.

$$(1 - b)^k / (1 - b + c p \cdot d - c p \cdot z)^k, \quad 1 > b > c p \cdot d > 0.$$

Then the amount of probability at the origin is

$$(1 - b)^k / (1 - b + c\tilde{p}\tilde{d})^k > (1 - b)^k.$$

Suppose now that the probability at the origin is reduced by an amount $(1 - b)^k$, before k is allowed to tend to zero in the usual Fisherian manner. The resultant distribution is then MLSD-m, since

$$\begin{aligned} & \lim_{k \rightarrow 0} \left[\frac{\left(\frac{1 - b}{1 - b + c\tilde{p}\tilde{d} - c\tilde{p}\tilde{z}} \right)^k - (1 - b)^k}{1 - (1 - b)^k} \right] \\ &= \lim_{k \rightarrow 0} \left[\frac{(1 - b + c\tilde{p}\tilde{d} - c\tilde{p}\tilde{z})^{-k} - 1}{(1 - b)^k - 1} \right] \\ &= \log(1 - b + c\tilde{p}\tilde{d} - c\tilde{p}\tilde{z}) / \log(1 - b). \end{aligned}$$

Consider also BCD-m with p.g.f.

$$\{\theta + (1 - \theta)\tilde{p}\tilde{z}\} / \{1 + \phi - \phi\tilde{p}\tilde{z}\}, \quad 1 < \theta < 1 + \phi.$$

Following the univariate treatment by Steutel (1968) and Kemp (1979), this can be shown to be infinitely divisible. Hence a k -fold convolution of such distributions will yield a valid distribution even when k is positive but not an integer. It is natural to ask whether LSD-m or MLSD-m can arise from it by a Fisherian-type limiting process. This does not occur; nevertheless the resultant distribution is closely related to MLSD-m. We have, remembering that $\tilde{p}\tilde{d} = 1$,

$$\begin{aligned} & (\theta + (1 - \theta)\tilde{p}\tilde{z}) / (1 + \phi - \phi\tilde{p}\tilde{z}) \\ &= 1 - A + A / (1 + \phi - \phi\tilde{p}\tilde{z}) \quad \text{where } A = (1 - \theta + \phi) / \phi. \end{aligned}$$

And so, taking $0 < C < \theta / (1 + \phi)$,

$$\begin{aligned} & \lim_{k \rightarrow 0} \left[\frac{[1 - A + A / (1 + \phi - \phi\tilde{p}\tilde{z})]^k - C^k}{1 - C^k} \right] \\ &= \log \left\{ \frac{1 - A}{C} + \frac{A}{C} (1 + \phi - \phi\tilde{p}\tilde{z}) \right\} / \log(1/C) \\ &= \frac{\log\{\theta - (\theta - 1)\tilde{p}\tilde{z}\} - \log(1 + \phi - \phi\tilde{p}\tilde{z}) - \log C}{- \log C} \end{aligned}$$

If $C = \theta/(1 + \phi)$, i.e. all the probability at the origin is truncated, then the resultant distribution has p.g.f.

$$\frac{-\log\{1 - \phi \tilde{p} \cdot \tilde{z}/(1 + \phi)\} + \log\{1 - (\theta - 1)\tilde{p} \cdot \tilde{z}/\theta\}}{-\log\{1 - \phi/(1 + \phi)\} + \log\{1 - (\theta - 1)/\theta\}}.$$

Note that $(\theta - 1)/\theta < \phi/(1 + \phi)$, since $\theta \leq 1 + \phi$.

These limit results involving MLSD-m appear to be completely new.

3.7 Mixed shifted-geometric model. Let us return to Kendall's mixed birth-and-death process given in Section 3.3. When $\beta = \mu$, the birth-and-death process gives rise to an unmodified logarithmic distribution only when $\beta T = \log(1 + \beta T)$. This is a very artificial restriction. However, when $\mu = 0$, the initial distribution is always shifted geometric and the resultant distribution is always LSD-1, see Kemp's (1981) algorithm BM for generating pseudo-random variables from LSD-1. The multivariate analogue of this univariate model is

$$\int_0^\alpha \frac{\tilde{p} \cdot \tilde{z}(1 - v)}{1 - v\tilde{p} \cdot \tilde{z}} \frac{dv}{(v - 1)\log(1 - \alpha)} = \frac{\log(1 - \alpha\tilde{p} \cdot \tilde{z})}{\log(1 - \alpha\tilde{p} \cdot \tilde{d})}$$

where $\tilde{p} \cdot \tilde{d} = 1$.

3.8 Mixed BCD model. Suppose now that the mixing distribution with p.d.f.

$$1/(1 + u)\log(1 + L), \quad 0 \leq u \leq L$$

be applied to the BCD-m with p.g.f.

$$(1 - \phi + \phi\tilde{p} \cdot \tilde{z})/(1 + u\phi - u\phi\tilde{p} \cdot \tilde{z}), \quad \tilde{p} \cdot \tilde{d} = 1.$$

The resultant distribution is MLSD-m since its p.g.f. is

$$\left[\frac{\log\{1 - u/(1 + u) + \phi u(1 - \tilde{p} \cdot \tilde{z})/(1 + u)\}}{-\log\{1 + L\}} \right]_0^L \\ = [\log\{1 + L\} - \log\{1 + \phi L - \phi L\tilde{p} \cdot \tilde{z}\}]/\log\{1 + L\}.$$

3.9 Mixed shifted-Poisson model. The mixed shifted-geometric model in subsection 3.7 has important implications. Since the

geometric is a mixed Poisson distribution, LSD-1 must be a mixed shifted-Poisson. We have

$$\int_0^{\infty} s e^{\lambda(s-1)} e^{-\lambda/u} (1/u) d\lambda = s/(1+u-us);$$

because repeated mixing operations obey the associative law, the required mixing distribution therefore has the p.d.f.

$$\int_0^L \frac{e^{-\lambda/u}}{u} \cdot \frac{du}{(1+u)\log(1+L)} = \frac{e^{\lambda} E_1(\lambda + \lambda/L)}{\log(1+L)}$$

where λ has support $[0, \infty)$ and $E(\cdot)$ is the exponential integral. Finally we obtain

$$\int_0^{\infty} s e^{\lambda(s-1)} \frac{e^{\lambda} E_1(\lambda + \lambda/L)}{\log(1+L)} d\lambda = \frac{\log\{1 - Ls/(1+L)\}}{\log\{1 - L/(1+L)\}}.$$

Thus the shifted univariate logarithmic distribution is a mixed Poisson. The multivariate analogue of the model is

$$\begin{aligned} & \int_0^{\infty} (\tilde{p} \cdot \tilde{z}) e^{\lambda(\tilde{p} \cdot \tilde{z} - 1)} \frac{e^{\lambda} E_1(\lambda + \lambda/L)}{\log(1+L)} d\lambda \\ & = \log\{1 - L\tilde{p} \cdot \tilde{z}/(1+L)\} / \log\{1 - L/(1+L)\} \end{aligned}$$

where $\tilde{p} \cdot \tilde{d} = 1$.

3.10 The Poisson cluster model for LSD-1. Katti (1967) proved the infinite divisibility of LSD-1 when shifted so as to have support $0, 1, 2, \dots$. The shifted LSD-1 therefore arises from a Poisson distribution of clusters of varying sizes. Explicit formulae for the cluster-size probabilities have been given by Kemp (1978). The multivariate analogue of this property of LSD-1 is

$$\exp[h(\tilde{p} \cdot \tilde{d})\{h(\tilde{p} \cdot \tilde{z})/h(\tilde{p} \cdot \tilde{d}) - 1\}] = \log(1 - \alpha \tilde{p} \cdot \tilde{z}) / [\tilde{p} \cdot \tilde{z} \log(1 - \alpha)]$$

where $h(s)/h(1)$ is the univariate cluster size distribution,

$$h(u) = \sum_{r \geq 1} B_r^{(r)} (-u)^r / r! r, \quad \tilde{p} \cdot \tilde{d} = 1, \quad \text{and} \quad B_r^{(n)}(\cdot) \text{ is the general-}$$

ized Bernoulli polynomial. The infinite divisibility property of shifted LSD-1 cannot be extended to LSD-m.

3.11 Multivariate damage models. Multinomial classification of univariate counted objects creates a multivariate distribution

from a univariate one; this is the multivariate generalization of the Rao damage model. Classification using SMD-m converts LSD-1 into LSD-m; classification using MD-m converts LSD-1 into MLSD-m.

All the modes of genesis in this section have been given in their multivariate forms for the sake of generality. The focus of interest in this paper is however LSD-2; bivariate versions of all the above models can be obtained by taking $m = 2$. The importance of LSD-2 in many areas of application becomes clear when one considers the wide scope of the models given in this section.

4. CONDITIONALITY PROPERTIES OF HOMOGENEOUS BIVARIATE DISTRIBUTIONS

Consider now any homogeneous bivariate distribution with p.g.f. $G_{x,y}(s,t) = H(as + bt)$. Then

$$G_{x|y}(s) = G_{x,y}(s,1) = H(as + b)$$

$$G_{x+y}(s) = G_{x,y}(s,s) = H(as + bs)$$

$$G_{x-y}(s) = G_{x,y}(s,1/s) = H(as + b/s).$$

Subrahmaniam (1966) has shown that for non-homogeneous as well as homogeneous bivariate distributions

$$G_{x|y=n}(s) = \left[\frac{\partial^n G_{x,y}(s,t)}{\partial t^n} \right]_{s=1} \bigg/ \left[\frac{\partial^n G_{x,y}(s,t)}{\partial t^n} \right]_{t=0}$$

A simpler result can be obtained when (X, Y) has a homogeneous bivariate distribution, for then $G_{x,y}(s,t) = H(as + bt)$ and $G_{x|y=n}(s) = H^{(n)}(as)/H^{(n)}(a)$. Note that this is independent of b .

Using a Subrahmaniam type of argument, the distribution of X conditional on $X + Y = n$ can be shown to have the SMD-1 p.g.f.

$$G_{x|x+y=n}(s) = \left[\frac{\partial^n H(as + bt)}{\partial t^n} \bigg/ \left[\frac{\partial^n H(as + bt)}{\partial t^n} \right]_{s=1} \right]_{t=0} = \left[\frac{as + b}{a + b} \right]^n,$$

whatever the form of $H(\cdot)$.

For the distribution of X conditional on $X - Y = n$ we can show that

$$G_{x|y-x=n}(s) = \frac{\text{coeff. of } t^n \text{ in } H(as/t + bt)}{\text{coeff. of } t^n \text{ in } H(a/t + bt)}.$$

Now $H(as/t + bt)$ will have a Laurent expansion in t . If it is possible to separate it into the sum of two functions H_1 and H_2 with power series expansions in t and $1/t$, respectively, then the conditional p.g.f.'s can again be obtained by partial differentiation.

Consider now the p.g.f. for LSD-2

$$G_{x,y}(s,t) = \log(1 - as - bt)/\log(1 - a - b).$$

The two marginal distributions are both MLSD-1, with p.g.f.'s

$$G_x(s) = \log(1 - as - b)/\log(1 - a - b)$$

and $G_y(s) = \log(1 - a - bt)/\log(1 - a - b).$

The marginal distribution of $X + Y$ is LSD-1, with p.g.f.

$$G_{x+y}(s) = \log\{1 - (a + b)s\}/\log(1 - a - b).$$

The conditional distribution of $X|Y = 0$ is LSD-1, with p.g.f.

$$G_{x|y=0}(s) = \log(1 - as)/\log(1 - a).$$

Similarly the conditional distribution of $Y|X = 0$ is LSD-1 with p.g.f.

$$G_{y|x=0}(t) = \log(1 - bt)/\log(1 - b).$$

However, the conditional distributions of $Y|X = n$ and $X|Y = n$, $n \neq 0$, are NBD-1, not LSD-1; the p.g.f.'s are

$$G_{x|y=n}(s) = \left(\frac{1-a}{1-as}\right)^n \quad \text{and} \quad G_{y|x=n}(t) = \left(\frac{1-b}{1-bt}\right)^n.$$

The conditional distribution of $X|X + Y = n$ is necessarily SMD-1. We have

$$G_{x|x+y=n} = \left(\frac{as+b}{a+b}\right)^n \quad \text{and} \quad G_{y|x+y=n} = \left(\frac{a+bs}{a+b}\right)^n$$

All these results are special cases of the conditionality results given by Patil and Bildikar (1967), and can of course also be obtained as special cases of the results for homogeneous bivariate distributions.

We come now to new results; these concern the marginal and conditional distributions relating to the difference between the values of X and Y . The marginal distribution of $X - Y$ has the p.g.f.

$$\begin{aligned} G_{X-Y}(s) &= \frac{\log(1 - as - b/s)}{\log(1 - a - b)} \\ &= \frac{\log[(1 - a\lambda s)(1 - b\lambda/s)/\lambda]}{\log(1 - a - b)} \\ &= \frac{\log(1 - a\lambda s) + \log(1 - b\lambda/s) - \log(\lambda)}{\log(1 - a - b)} \end{aligned}$$

where $\lambda = \{1 - (1 - 4ab)^{1/2}\}/2ab$. Hence

$$\text{Prob}(x - y = 0) = -\log(\lambda)/\log(1 - a - b).$$

Moreover $(X - Y)|X > Y$ is LSD-1 with p.g.f. given by $\log(1 - a\lambda s)/\log(1 - a\lambda)$. Similarly $(Y - X)|X < Y$ is LSD-1 with p.g.f. $\log(1 - b\lambda s)/\log(1 - b\lambda)$.

Consider now the distribution of $X|Y - X = n$, $n \neq 0$. We have

$$\begin{aligned} G_{X|Y-X=n}(s) &= \frac{\text{coeff. of } t^n \text{ in } \log(1 - as/t - bt)}{\text{coeff. of } t^n \text{ in } \log(1 - a/t - bt)} \\ &= \frac{\text{coeff. of } t^n \text{ in } \log(1 - b\eta t)}{\text{coeff. of } t^n \text{ in } \log(1 - b\lambda t)}, \end{aligned}$$

where $\eta = \{1 - (1 - 4abs)^{1/2}\}/2abs$. Hence

$$\begin{aligned} G_{X|Y-X=n}(s) &= \left[\frac{1 - (1 - 4abs)^{1/2}}{2as} \right]^n \left/ \left[\frac{1 - (1 - 4ab)^{1/2}}{2a} \right]^n \right. \\ &= \frac{{}_2F_1[n/2, (n+1)/2; n+1; 4abs]}{{}_2F_1[n/2, (n+1)/2; n+1; 4ab]}, \end{aligned}$$

which we can recognize as the univariate lost-games distribution from Kemp and Kemp (1968). Notice that this is symmetric in (ab) ; consideration of the coefficients of t^{-n} in the requisite expressions establishes that

$$G_{x|x=y=n}(s) = s^n G_{x|y-x=n}(s).$$

$$\begin{aligned} \text{Finally } G_{x|x=y}(s) &= \frac{\log(1 + asb\eta^2)}{\log(1 + ab\lambda^2)} \\ &= \frac{\log[1 + \{1 - (1 - 4abs)^{\frac{1}{2}}\}^2/4abs]}{\log[1 + \{1 - (1 - 4ab)^{\frac{1}{2}}\}^2/4ab]} \\ &= \log \left[\frac{1 - (1 - 4abs)^{\frac{1}{2}}}{2abs} \right] / \log \left[\frac{1 - (1 - 4ab)^{\frac{1}{2}}}{2ab} \right] \\ &= \frac{{}_3F_2[3/2, 1, 1; 2, 2; 4abs]}{{}_3F_2[3/2, 1, 1; 2, 2; 4ab]}, \end{aligned}$$

see Kemp and Kemp (1969), Kemp (1978). This is the univariate distribution which gives rise to the lost-games distribution when used as a Poisson cluster-size distribution. Note that it is a distribution with support $1, 2, \dots$.

5. CHI-SQUARE GOODNESS-OF-FIT APPLICATION

The usefulness of these marginality and conditionality results becomes clear when we consider the chi-square goodness-of-fit test for examining whether or not data generated pseudo-randomly from an LSD-2 are fitted by the distribution. Four sets of data, each comprising 120 data points, were generated from the LSD-2 with $p = (0.1, 0.8)$.

The usual chi-square procedure is to evaluate the expected frequencies, grouping them so as to avoid very low expected frequencies. The chi-square test is of course only approximate and there is disagreement concerning the lowest 'safe' size for an expected frequency; with a desire to be noncontroversial we have adopted a minimum expected-frequency size of 3. For a sample size of 120 this corresponds to minimum expected probability equal to 0.025. Note that because the data have been generated by a pseudo-random process, see Kemp (1980), the theoretical values of the parameters are known. More usually it would be necessary to estimate their values - Patil and Bildikar (1967) show how maximum likelihood estimates can be obtained.

Problems arise as soon as we attempt to group the small expected frequencies. It is customary to consider these either row-by-row or column-by-column in the two-way array of expected frequencies. Sometimes this leads to odd-looking groupings. (Of

course the grouping procedure should be independent of the observed frequencies.)

In order to avoid giving precedence to either rows or columns, let us consider the diagonals given by $X - Y = n$. For LSD-2 with $a = 0.1$, $b = 0.8$, we find that $\lambda = 1.096$.

$$\text{Since } P(x - y = n) = -a^n \lambda^n / n \log(1 - a - b), \quad n > 0,$$

$$P(x - y = 0) = -\log(\lambda) / \log(1 - a - b),$$

$$P(x - y = n) = -b^n \lambda^n / n \log(1 - a - b), \quad n < 0,$$

we calculate

$$\begin{array}{ll} P(x - y \geq 2) = 0.003 & P(x - y = -6) = 0.033 \\ P(x - y = 1) = 0.048 & P(x - y = -7) = 0.024 \\ P(x - y = 0) = 0.040 & P(x - y = -8) = 0.019 \\ P(x - y = -1) = 0.381 & P(x - y = -9) = 0.015 \\ P(x - y = -2) = 0.167 & P(x - y = -10) = 0.012 \\ P(x - y = -3) = 0.098 & P(x - y = -11) = 0.009 \\ P(x - y = -4) = 0.064 & P(x - y \leq -12) = 0.043. \\ P(x - y = -5) = 0.045 \end{array}$$

Note that a logarithmic series is fitted on each side of the origin; when the probabilities get smaller than the somewhat arbitrarily chosen figure of 0.025, the standard method of grouping for a univariate distribution can be used. These groupings are indicated by the brackets.

Consider next the ungrouped probabilities greater than 0.025×2 . This leads to an examination of the conditional distributions of $(X|X - Y = -1)$, $(X|X - Y = -2)$, $(X|X - Y = -3)$, $(X|X - Y = -4)$. The tails of these distributions can again be grouped in the usual manner, making sure that probabilities greater than 0.025 are attained everywhere.

Table 1 gives the final groupings in column (1), together with their expected frequencies in column (2). Columns (3)-(6) show the observed frequencies for the four data sets and the corresponding chi-square values.

TABLE 1: *Bivariate frequency distributions for four pseudo-randomly-generated data sets from LSD-2, a = 0.1, b = 0.8.*

Group	Expected frequency	Observed frequencies for four data sets			
$x > y$	6.05	7	7	6	6
$x = y$	4.78	4	7	5	6
$x = 0, y = 1$	41.69	36	42	41	46
$x > 0, y = x + 1$	4.01	6	5	5	6
$x = 0, y = 2$	16.68	13	13	21	18
$x > 0, y = x + 2$	3.36	4	5	2	0
$y = x + 3$	11.71	19	13	9	11
$y = x + 4$	7.70	10	7	6	5
$y = x + 5$	5.40	7	2	7	4
$y = x + 6$	3.95	2	3	6	2
$y - x = 7, 8$	5.25	3	3	3	7
$y - x = 9, 10, 11$	4.29	5	8	3	1
$y - x \leq 12$	5.12	4	5	6	8
$\chi^2_{(12)}$		10.95	9.78	5.98	12.25

The grouping procedure uses the results obtained in Section 4 concerning the bilateral marginals and their associated conditional distributions for LSD-2. It has been adopted because it is an objective analogue of the usual univariate procedure. As in the univariate case the only real opening for controversy concerns the choice of minimum expected-frequency size.

REFERENCES

- Bates, G. E. and Neyman, J. (1952). Contributions to the theory of accident proneness. *University of California Publications in Statistics*, 1, 215-275.
- Boswell, M. T. and Patil, G. P. (1971). Chance mechanisms generating the logarithmic series distribution used in the analysis of number of species and individuals. In *Statistical Ecology, Vol. 1*, G. P. Patil *et al.*, eds. The Pennsylvania State University Press, University Park. Pages 99-130.
- Chatfield, C., Ehrenberg, A. S. C. and Goodhardt, G. J. (1966). Progress on a simplified model of stationary purchasing behaviour. *Journal of the Royal Statistical Society, Series A*, 129, 317-367.
- Edwards, C. B. and Gurland, J. (1961). A class of distributions applicable to accidents. *Journal of the American Statistical Association*, 56, 503-517.

- Fisher, R. A., Corbet, A. S. and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample from an animal population. *Journal of Animal Ecology*, 12, 42-58.
- Guldberg, A. (1934). On discontinuous frequency functions of two variables. *Skandinavisk Aktuarietidskrift*, 17, 89-117.
- Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*. Houghton Mifflin, Boston.
- Katti, S. K. (1967). Infinite divisibility of integer-valued random variables. *Annals of Mathematical Statistics*, 38, 1306-1308.
- Kemp, A. W. (1978). Cluster size probabilities for generalized Poisson distributions. *Communications in Statistics, Series A*, 7, 1433-1438.
- Kemp, A. W. (1979). Convolutions involving binomial pseudo-variables. *Sankhyā, Series A*, 41, 232-243.
- Kemp, A. W. (1981). Frugal methods of generating bivariate discrete random variables. In *Statistical Distributions in Scientific Work*, C. Taillie, G. P. Patil, and B. Baldessari, eds. Reidel, Dordrecht-Holland. Vol. 4, 321-329.
- Kemp, A. W. (1981). Efficient generation of the logarithmic distribution. *Applied Statistics* (in press).
- Kemp, A. W. and Kemp, C. D. (1968). On a distribution associated with certain stochastic processes. *Journal of the Royal Statistical Society, Series B*, 30, 160-163.
- Kemp, A. W. and Kemp, C. D. (1969). Branching and clustering models associated with the 'lost-games' distribution. *Journal of Applied Probability*, 6, 700-703.
- Kendall, D. G. (1948). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, 35, 6-15.
- Khatrı, C. G. (1959). On certain properties of power series distributions. *Biometrika*, 46, 486-490.
- McCloskey, J. W. (1965). A model for the distribution of individuals by species in an environment. Technical Report No. 7, Department of Statistics, Michigan State University.
- Nelson, W. C. and David, H. A. (1967). The logarithmic distribution: a review. *Virginia Journal of Science*, 18, 95-102.
- Neyman, J. (1965). Certain chance mechanisms involving discrete distributions. In *Classical and Contagious Discrete Distributions*, G. P. Patil, ed. Pergamon Press, London. Pages 4-14.
- Patil, G. P. and Bildikar, S. (1967). Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties. *Journal of the American Statistical Association*, 62, 655-674.
- Patil, G. P. and Joshi, S. W. (1968). *A Dictionary and Bibliography of Discrete Distributions*. Oliver and Boyd, Edinburgh.
- Philippou, N. and Roussas, G. G. (1974). A note on the multivariate logarithmic series distribution. *Communications in Statistics*, 3, 469-472.

- Quenouille, M. H. (1949). A relation between the logarithmic, Poisson and negative binomial series. *Biometrics*, 5, 162-164.
- Sibuya, M., Yoshimura, I. and Shimizu, R. (1964). Negative multinomial distribution. *Annals of the Institute of Statistical Mathematics Tokyo*, 26, 409-426.
- Steutel, F. W. (1968). A class of infinitely divisible mixtures. *Annals of Mathematical Statistics*, 39, 1153-1157.
- Subrahmanian, K. (1966). A test for "intrinsic correlation" in the theory of accident proneness. *Journal of the Royal Statistical Society, Series B*, 28, 180-189.

[Received July 1980. Revised October 1980]

A SURVEY OF ESTIMATING DISTRIBUTIONAL PARAMETERS AND SAMPLE SIZES FROM TRUNCATED SAMPLES

SAUL BLUMENTHAL

Department of Mathematics
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801 USA

SUMMARY. Suppose X_1, \dots, X_N are independent random variables with common distribution $F(x; \theta)$. We observe X_i only if it lies outside a given region R . Thus the number n of observed X 's is a binomial (N, P) variate, where $P = 1 - P(X \text{ in } R)$. Based on n and the values of the observed X 's, we want to estimate N and θ . Conditional and unconditional maximum likelihood estimators and modifications of these will be discussed. Asymptotic results of both first and second order will be considered, and small sample results also mentioned. Sequential and two stage results will receive brief coverage. A variety of applications will be given.

KEY WORDS. Estimating n , sample size estimation, asymptotic expansions, second order efficiency, truncated samples.

1. DISCUSSION OF RESULTS

In many sampling situations, observations are restricted to a part of the possible range of population values. Of a set of N potential observations, only n remain observable as a result of the restricting process, while $(N-n)$ are eliminated by it. The resulting set of incomplete data is commonly referred to as censored when N (or $N-n$) is known, and as truncated otherwise. In life testing which is designed to estimate the average life time of, for example, transistors from a given production line, typically, a certain number of items are tested for a fixed length of time. This gives rise to a censored sample of lifetimes since the lifetimes of items surviving the life test are unknown.

In certain life testing situations the total number of items on test is not known. This happens, for instance, when among the items put on a life test there is a certain unknown number of items with a particular defect identifiable only after the item fails. If the lifetime of an item with this particular defect is the variable of interest, then, the sample is truncated in that the number of missing observations of lifetime greater than the burn-in or testing period is unknown. An important problem that arises in such situations is estimating the number of remaining defectives of a particular type after an initial burn-in period (Blumenthal and Marcus, 1975a,b). Sequential estimation procedures in which the test period is not pre-determined and for which one can guarantee with a high probability that the observed number of defectives of the given type equals the total number in the batch are of special interest in reliability. They can be applied as screening or sequential burn-in procedures which, with a specified level of confidence, remove all defectives in the batch (Marcus and Blumenthal, 1974). In the quality control context in which a defective can be identified without a life test, such screening procedures have been in use for many years.

Censored and truncated samples also arise in production engineering when sorting or inspection procedures eliminate items above or below designated tolerance limits. The measurements made on the items within the tolerance limits represent either a censored or truncated sample depending on whether a count was kept of the number which failed to meet the specifications.

Another interesting example arises in the software reliability literature. Jelinski and Moranda (1972) postulate a model for the initial error content N of a program. This model can be paraphrased by saying that with each error is associated a random variable representing the running time of the program up to the time of detection of the error. These errors are independent identically distributed random variables. After running the program for a fixed period of time, an estimate of the number of undetected errors is required. Technically, this is exactly the same problem as estimating the number of remaining defectives after a life test.

Many sociological, epidemiological and genetic or medical investigations seek to study a target population of persons characterized by a trait that occurs only rarely in the population at large. Ascertainment of the study group is often done by merging preexisting but incomplete lists (such as birth records, death certificates, records from hospitals, schools and other institutions and disease registers) of members of the target population. However, there may be cases not listed by any source. Thus, the problem is to estimate the number of

such unlisted cases from a truncated sample (Wittes, 1970; Wittes *et al.*, 1974). A similar situation arises in visual scanning experiments in particle physics where film containing photographs of particles is scanned separately by two or more scanners all of whom might miss some particles with very poor visibility (Sanathanan, 1972b).

The truncated Poisson typically arises when the zero class is unobservable, owing to the fact that only cases where at least one event occurs are reported while the total number of cases is unknown. An example would be the number of accidents per worker in a factory. While it is a simple matter to count over a given period of time the number of workers sustaining one or more accidents, the number of persons who did not have an accident could not be enumerated owing to the factory population fluctuating during that period. Another type of Poisson data with the zero class missing would be that of the number of persons per house who are infected with a disease such as measles. The zero class represents households of which one or more member has come in contact with an infected individual, but none of its members has contracted the disease. The number of such households is difficult to assess.

An example of the truncated negative binomial distribution is the following, given by Sampford (1955). If chromosome breaks in irradiated tissue can occur only in those cells which are at a particular stage of the mitotic cycle at the time of irradiation, a cell can be demonstrated to have been at that stage only if breaks actually occur. Thus in the distribution of breaks per cell, cells not susceptible to breakage are indistinguishable from susceptible cells in which no breaks occur.

Goel and Joglekar (1976) fit the truncated negative binomial distribution to field failure data for several types of systems. For each system type, N units were in use for a known period of time T , and a count n_x was available showing how many of the units failed x times ($x = 1, 2, \dots$) during the test period. No count was available for the number of units showing zero failures in the test period, i.e. N and n_0 were unknown.

A somewhat different example is given by Rao *et al.* (1973) who fit a negative binomial distribution to the counts of number of children per family in a given sample. In this sample there are M childless families which are believed to consist of two subgroups: one which is basically sterile and biologically a separate group, and one which is fertile but did not have children. The number n_0 in the latter group is the required frequency, and except for being bounded above by M , it is

unknown. Thus, the zero-truncated negative binomial distribution must be used.

Censoring or truncation of a random variable X implies that there is a region R where if X is in R , X is not observable. In the case of a continuous variable (e.g. survival or reaction time) this generally means the values of the missing observations are known to lie in a certain open ended interval. On the other hand, for a discrete variable with a missing category it means that the missing observations belong to some known category, although their number is unknown. Such data are therefore necessarily of the truncated type.

There are three distinct sampling rules that are associated with truncated data.

1. The total number of random variables N is fixed, but unknown. The n observations obtained represent the values among the N which fall inside the observable region \bar{R} , (\bar{R} is the complement of R). In this case, n is a binomial random variable with parameters N and $P = P(X \text{ in } \bar{R})$. This is referred to as binomial sampling, and will be the only situation covered in the present survey in detail.
2. N fixed and unknown as in 1. The n observations obtained represent the central order statistics with the smallest s_0 and largest s_1 values being censored. If s_0 and s_1 were known, this would be straightforward censored sampling. With s_0 and s_1 unknown, this could be referred to as pseudo-censored sampling, and has been discussed by Blumenthal and Marcus (1975a,b) and Johnson (1962, 1967, 1970, 1971, 1972, 1973, 1974a,b, 1976).
3. The underlying population from which sampling occurs is assumed to be infinite and values of the random variables in question are unobservable in a certain region R . Sampling is continued until a predetermined number n of observations is reached. The total number of random variables N generated in the process is now random with a Pascal distribution and is not known. This is referred to as the inverse binomial sampling rule. For further discussion of this situation, see Blumenthal and Sanathanan (1980).

Two types of problems are dealt with primarily in connection with the binomial sampling model for truncated data. One type calls for the estimation of the unknown number of missing observations. The software reliability problem, and one of the life testing problems discussed earlier are of this type. Estimation of the other parameters is only incidental in such problems whereas for the other type of truncated sampling problems although the number of missing observations is unknown, the focus is on the estimation of distributional parameters. This is the case with many life testing situations.

Estimation on the basis of truncated samples was dealt with in the early literature using the conditional approach almost exclusively. This approach consists of eliminating the unknown total sample size from consideration by assuming the number of observations to be fixed and then examining the conditional distribution of the given observations namely, the truncated distribution. The traditional approach to parameter estimation from truncated samples has been through these distributions. In many cases, the only estimation techniques available are of this conditional type. Thus, inference for truncated distributions plays an important role in any examination of inference for truncated samples. In particular, results of Tukey (1949) and Smith (1957) showing that sufficiency and completeness are inherited by the truncated distribution from the parent are particularly important. These have been applied to derive uniformly minimum variance unbiased estimators of reliability via the Rao-Blackwell technique by Holla (1967) for exponential distributions and Nath (1975) for gamma distributions. Carry over of monotonicity properties from the parent distribution can be helpful in developing moment estimators. Gross (1971) worked on monotonicity for truncated distributions.

A great many papers have appeared dealing with estimation of parameters of specific truncated distributions such as normal, exponential, Poisson, and negative binomial among others. For a thorough survey of the literature the books by Johnson and Kotz (1969, 1970a,b, 1972) are recommended.

Concern for estimating the number of missing observations appears much less often in the literature. When it has appeared, it has usually taken the form of a conditional estimator. If N is the total number of possible random variables and n is the actual number observed, we can estimate N by

$$\hat{N} = [n/\hat{P}] \quad (1)$$

where $[x]$ is the greatest integer in x and where \hat{P} is an estimator of $P(X \text{ in } \bar{R})$. The conditional approach is to estimate P through the truncated distribution. This approach

is found for instance in Hartley (1958) for the missing zero class of the negative binomial, in Dahiya and Gross (1973) for the Poisson, and in Sanathanan (1972a) for the multinomial.

Working with the exponential distribution, Blumenthal and Marcus (1975b) introduced the unconditional approach to estimating N and the distributional parameters. The unconditional likelihood is just the product of the conditional likelihood given n and the binomial likelihood for n . It thus takes the random nature of n into account and unconditional maximum likelihood estimates N can be appreciably different from their conditional counterparts. In this paper, it was observed that certain sample configurations cause the conditional \hat{N} to be undefined and a subset of these samples still leave the unconditional \hat{N} undefined. To avoid this situation, Blumenthal and Marcus examined Bayes modal estimators with respect to a conjugate prior density for the parameter of the exponential distribution. They found that these "modified" estimators had a form very similar to that of the maximum likelihood estimators and would not become undefined for any samples. In the sense of the usual asymptotic distribution theory, these estimators could not be distinguished from the m.l.e.'s, i.e. we could express \hat{N} as

$$\hat{N} = N + \sigma\sqrt{N} Z + O_p(1) \quad (2)$$

where Z is a function of the observations having mean zero and variance one whose distribution becomes normal as N increases, and the same Z and σ can be used to describe all of the estimators discussed so far. Thus, to make asymptotic distinctions among the estimators which might have some bearing on small sample properties, the $O_p(1)$ term was examined in more detail, being split into a constant β and a $O_p(1/\sqrt{N})$ term. It was found that the β term does differ among the estimators with the conditional estimator having a systematically larger β than its competitors, reflecting a tendency to overestimate even when finite valued. The expectation of β can be thought of as an "asymptotic bias." By manipulating the parameters of the conjugate weight function, they found a β which minimizes the maximum $E\beta$ over all parameter values. This particular "minimax bias" modified m.l.e. is a strong candidate to replace the m.l.e., and simulations by Watson and Blumenthal (1980) verify its superiority to the two competing m.l.e.'s.

This same approach was applied to estimating the missing zero class for the Poisson by Blumenthal, Dahiya and Gross (1978). Modified u.m.l.e.'s were shown to be superior to their unmodified competitors both asymptotically and for small samples.

For any single parameter distribution satisfying some mild regularity conditions, Blumenthal (1977) derives expressions for the "asymptotic bias" of N to allow comparisons of the various likelihood type estimators. The results are specialized to scale parameter families and the "minimax bias" modified m.l.e.'s are characterized.

For distributions depending on a multidimensional parameter θ , only first order asymptotic properties have been studied. Sanathanan has obtained the limiting joint normal distribution of N and $\hat{\theta}$ for the multinomial distribution (1972a) and for a more general class of regular distributions (1977).

If instead of trying to control asymptotic bias, one looks at a correction term to $E(\hat{N} - N)^2$, then the $O_p(1)$ term in (2) must be broken into $\beta + (\gamma/\sqrt{N}) + O_p(1/N)$ and the contribution of γ to the mean squared error examined. Watson and Blumenthal (1980) have done this and found modified estimators which approximately minimize maximum mean squared error. Monte Carlo comparisons suggest that these offer some improvement over the minimax bias estimators, but relatively little.

Blumenthal (1980) has considered both modified u.m.l.e.'s and modified c.m.l.e.'s obtained by using a conjugate prior on the conditional likelihood. The conjugate priors for the conditional and unconditional cases are not the same. For estimating N or a parameter θ , using minimax bias as a criterion, sometimes the modified u.m.l.e. is best and sometimes the modified c.m.l.e. Generally, modified c.m.l.e.'s share the desirable property that \hat{N} will exist for all samples. For obtaining first and second order asymptotic properties, the modified m.l.e.'s can be regarded as members of a class of estimators of the form $\hat{\theta} + (\rho(\hat{\theta})/n)$ where appropriate choice of $\rho(\theta)$ allows duplication of the asymptotic properties of any given modified m.l.e. starting with any modified m.l.e. $\hat{\theta}$. However, small sample properties are not readily reproduced, since $\hat{\theta} + (\rho(\hat{\theta})/n)$ will lead to nonexistent \hat{N} through (1), whenever $\hat{\theta}$ does. It is fairly common to use $\rho(\theta) = -nE(\hat{\theta} - \theta)$ so as to make the "adjusted" $\hat{\theta}$ unbiased to a second order approximation. Adjustment of \hat{N} can be achieved either through adjusting $\hat{\theta}$ in (1) or by directly creating $\hat{N} - \delta(\hat{\theta})$, where for instance $\delta(\theta) = E\beta$ might be used.

Extending results already available for complete samples, Blumenthal (1980) shows that the correction to mean squared error will in fact depend only on β under mild regularity conditions making derivation of the correction much simpler.

Relations between β and various measures of second order efficiency are explored, and Edgeworth expansions for the distribution of N are obtained and related to the expansion (2). Also conditional expansions of the form

$$N = (n/P) + A\sqrt{n} + B + C/\sqrt{n} + \dots \quad (3)$$

are examined. Using (3), expressions for $E[\hat{N} - (n/P)|n]$ and $E[(\hat{N} - (n/P))^2|n]$ can be obtained and these depend only on the truncated distributions. It is shown that comparisons between estimators on the basis of these criteria (to first or second order) are unchanged by obtaining the unconditional bias or mean squared error. Hence unconditionally valid comparisons among estimators may be made strictly on a conditional basis as had been done historically. However, this does not imply that the conditional estimators are necessarily as good as the unconditional ones with respect to the conditional comparisons.

A somewhat different approach to controlling mean squared error for estimating N is taken by Watson and Blumenthal (1981). For the exponential distribution, it was noted that σ in (2) is an unbounded increasing function of the ratio of the scaled parameter θ to the duration T of the test. For fixed duration tests, σ cannot be bounded. A two stage procedure is considered in which a fixed time first stage leads to an initial $\hat{\theta}$ which in turn is used to determine the duration of the second stage. It is shown that for large n , the total duration can be controlled so that for any θ , the resulting σ is below a specified value σ_0 .

In case the distribution is known completely so that only N is unknown, or if no inference regarding the distribution is desired, Blumenthal and Marcus (1975a) use Bayesian procedures which involve only a "marginal" prior for N .

When N is estimated, a goodness of fit test may be desired for some given density $f(x;\theta)$. Dahiya (1980b) has shown that the usual Pearson chi-squared test can be used but the result of estimating N is not to reduce the degrees of freedom by one. Instead, using \hat{N} causes the distribution to be that of a weighted average of two independent chi-squared variates which are somewhat larger than the single chi-square with one less degree of freedom.

Computing the simultaneous estimates of θ and N for modified unconditional m.l.e.'s can create some difficulties as an examination of Blumenthal and Marcus (1975b) or Blumenthal, Dahiya and Gross (1978) will show. Simplification of the computation can be achieved by using a technique developed by

Dahiya (1980a) for finding the modal estimator of any integer valued parameter.

2. APPLICATIONS

Many of the references cited contain real data from which an estimate of N has been computed. In most of the literature prior to Blumenthal and Marcus (1975b), only the c.m.l.e. was found. We have mentioned simulation studies to verify that the asymptotic theory reasonably describes small sample properties. In one paper, namely that of Jelinski and Moranda (1972), follow up studies did lead to a verification of the original N estimate. The authors assumed an exponential distribution, derived the u.m.l.e. assuming N to be continuous, and computed \hat{N} for three different data sets. The respective values of (n, \hat{N}) were (26, 31.1), (14, 17.4) and (12, 14.45), based on the production checkout phase. A subsequent test phase led to detection of additional errors and updated (n, \hat{N}) values of (31, 31.6), (15, 17.1) and (13, 14.54). Update data based on almost one more year of operation was reported by Moranda (1975) and showed n values of 32, 16, and 14 respectively. These can be taken as very close approximations to the true N . Using the formulae in Blumenthal and Marcus (1975b), we find that for the production phase data the integer valued u.m.l.e. and the minimum asymptotic bias modified u.m.l.e. agree with one another and take the values (31, 17, 14) respectively. The c.m.l.e.'s are computed to be (33, 22, 18) which are seen to be too large as asymptotic theory indicates.

ACKNOWLEDGEMENT

This research was supported in part by National Science Foundation Grant No. MCS 8002340, Department of Mathematics, University of Illinois, Urbana-Champaign.

REFERENCES

- Blumenthal, S. (1977). Estimating population size with truncated sampling. *Communications in Statistics*, A6, 297-308.
- Blumenthal, S. (1980). Stochastic expansions for point estimation from complete, censored and truncated samples. Submitted for publication.
- Blumenthal, S., Dahiya, R. C. and Gross, A. J. (1978). Estimating the complete sample size from an incomplete Poisson sample. *Journal of the American Statistical Association*, 73, 182-187.

- Blumenthal, S. and Marcus, R. (1975a). Estimating the number of items on life test. *IEEE Transactions on Reliability*, R-24, 271-273.
- Blumenthal, S. and Marcus, R. (1975b). Estimating population size with exponential failure. *Journal of the American Statistical Association*, 28, 913-922.
- Blumenthal, S. and Sanathanan, L. P. (1980). Estimation with truncated inverse binomial sampling. *Communications in Statistics*, A9, 997-1017.
- Dahiya, R. C. (1980a). An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician*, to appear.
- Dahiya, R. C. (1980b). Pearson goodness-of-fit test when the sample size is unknown. Unpublished manuscript.
- Dahiya, R. C. and Gross, A. J. (1973). Estimating the zero class from a truncated Poisson sample. *Journal of the American Statistical Association*, 68, 731-733.
- Goel, A. L. and Joglekar, A. M. (1976). Reliability acceptance sampling plans based upon prior distribution, III. Implications and determination of the prior distribution. Technical Report No. 76-3, Department of Industrial Engineering and Operations Research, Syracuse University.
- Gross, A. J. (1971). Monotonicity properties of the moments of truncated gamma and Weibull density functions. *Techometrics*, 13, 851-857.
- Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14, 174-194.
- Holla, M. S. (1967). Reliability estimation of the truncated exponential model. *Techometrics*, 9, 332-335.
- Jelinski, Z. and Moranda, P. (1972). Software reliability research. In *Statistical Computer Performance Evaluation*, W. Freiburger, ed. Academic Press, New York. Pages 465-484.
- Johnson, N. L. (1962). Estimation of sample size. *Technometrics*, 4, 59-67.
- Johnson, N. L. (1967). Sample censoring. ARO-D Rep. 67-2, U. S. Army Research Office - Durham, Durham, N.C.
- Johnson, N. L. (1970). A general purpose test of censoring of extreme sample values. In *Essays in Probability and Statistics* (S. N. Roy Memorial Volume), R. C. Bose, ed. University of North Carolina Press, Chapel Hill, Pages 377-384.
- Johnson, N. L. (1971). Comparison of some tests of sample censoring of extreme values. *Australian Journal of Statistics*, 13, 1-6.
- Johnson, N. L. (1972). Inferences on sample size: sequences of samples. *Trabajos de Estadística*, 23, 85-110.
- Johnson, N. L. (1973). Robustness of certain tests of censoring of extreme values. Institute of Statistics mimeo series No. 866, University of North Carolina, Chapel Hill.

- Johnson, N. L. (1974a). Robustness of certain tests of censoring of extreme values, II: some exact results for exponential populations. Institute of Statistics mimeo series No. 940, University of North Carolina, Chapel Hill.
- Johnson, N. L. (1974b). Estimation of rank order. Institute of Statistics mimeo series No. 931, University of North Carolina, Chapel Hill.
- Johnson, N. L. (1976). Completeness comparisons among sequences of samples. Institute of Statistics mimeo series No. 1056, University of North Carolina, Chapel Hill.
- Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*. Houghton Mifflin, Boston.
- Johnson, N. L. and Kotz, S. (1970a). *Continuous Univariate Distributions-1*. Houghton Mifflin, Boston.
- Johnson, N. L. and Kotz, S. (1970b). *Continuous Univariate Distributions-2*. Houghton Mifflin, Boston.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Marcus, R. and Blumenthal, S. (1974). A sequential screening procedure. *Techometrics*, 16, 229-234.
- Moranda, P. B. (1975). Prediction of software reliability during debugging. MDAC paper WD 2471, McDonnell Douglas Astronautics Company-West, Huntington Beach, California.
- Nath, G. B. (1975). Unbiased estimates of reliability for the truncated gamma distribution. *Scandinavian Actuarial Journal*, 1, 181-186.
- Rao, B. R., Mazumdar, S., Waller, J. H. and Li, C. C. (1973). Correlation between the numbers of two types of children in a family. *Biometrics*, 29, 271-279.
- Sampford, M. R. (1955). The truncated negative binomial distribution. *Biometrika*, 42, 58-69.
- Sanathanan, L. P. (1972a). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, 43, 142-152.
- Sanathanan, L. P. (1972b). Models and estimation methods in visual scanning experiments. *Technometrics*, 14, 813-840.
- Sanathanan, L. P. (1977). Estimating the size of a truncated sample. *Journal of the American Statistical Association*, 72, 669-672.
- Smith, W. L. (1957). A note on truncation and sufficient statistics. *Annals of Mathematical Statistics*, 28, 247-252.
- Tukey, J. W. (1949). Sufficiency, truncation, and selection. *Annals of Mathematical Statistics*, 20, 309-311.
- Watson, D. and Blumenthal, S. (1980). Estimating the size of a truncated sample. *Communications in Statistics*, A9, 1535-1550.
- Watson, D. and Blumenthal, S. (1981). A two stage procedure for estimating the size of a truncated exponential sample. *Australian Journal of Statistics*, 20, No. 3.

- Wittes, J. T. (1970). *Estimation of population size: the Bernoulli census*. Ph.D. thesis, Harvard University.
- Wittes, J. T., Cotton, T. and Sidel, V. W. (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *Journal of Chronic Disease*, 27, 25-36.

[Received June 1980. Revised October 1980]

PARAMETER ESTIMATION FOR TRUNCATED EXPONENTIAL FAMILIES

BRADFORD R. CRAIN

Department of Mathematics
Portland State University
Portland, Oregon 97207 USA

LOREN COBB

Department of Biometry
Medical University of South Carolina
Charleston, South Carolina 29403 USA

SUMMARY. The multiparameter exponential families of probability density functions include many commonly used densities such as the normal, gamma, and beta, and many others besides. We consider the general parameter estimation problem for these families, given observations that are restricted (i.e. truncated) to the closed interval $[a,b]$. We show how to find maximum likelihood estimators using Newton-Raphson iteration, but observe that in general this technique requires numerical integration within each iteration. Then we give a new non-iterative method for obtaining the desired estimators for the parameter vector. An example using the generalized Inverse Gaussian is given.

KEY WORDS. Exponential families, maximum likelihood, non-iterative estimation.

1. INTRODUCTION

Let X be a random variable which has probability density function $f(x)$ (with respect to Lebesgue measure), where we assume $f(x)$ to be that of a canonical exponential family of truncated densities, meaning that

■

$$f(x) = \exp \left[\sum_{i=1}^m \tau_i \gamma_i(x) - \eta(\tau) \right] \quad (1)$$

for $x \in [a, b]$ and zero otherwise. The functions $\gamma_i(x)$ are bounded on $[a, b]$, measurable and linearly independent. The parameter vector τ is an element of R^m , and the normalizing function $\eta(\tau)$ is determined by the condition that $\int_a^b f(x) dx = 1$. The class of densities encompassed by (1) include the exponential, normal, gamma, beta, inverse gamma, inverse Gaussian, and many more.

It is well known (see Barndorff-Nielsen, 1978, or Lehmann, 1959, for example) that canonical exponential densities of the form (1) have many special properties. For instance, it can be shown that

$$\partial \eta(\tau) / \partial \tau_j = E[\gamma_j(X)] \quad \text{and} \quad \partial^2 \eta(\tau) / \partial \tau_j \partial \tau_k = \text{Cov}[\gamma_j(X), \gamma_k(X)].$$

Let Δ be the gradient, and $\gamma(x) = \{\gamma_1(x), \dots, \gamma_m(x)\}$. Then

$$\Delta \eta(\tau) = E[\gamma(X)] \quad \text{and} \quad H(\tau) = \Delta^2 \eta(\tau) = \text{Cov}[\gamma(X), \gamma(X)].$$

Thus the Hessian matrix $H(\tau)$ of $\eta(\tau)$ is the covariance matrix of the random variables $\gamma_i(X)$, for $i = 1, \dots, m$. Let γ_0 be the function that is identically 1 on $[a, b]$. We will assume throughout that the functions $\{\gamma_i(x) : i = 0, \dots, m\}$ are linearly independent. Under this assumption $H(\tau)$ is positive definite for each $\tau \in R^m$.

Now consider the problem of maximum likelihood estimation for τ . Let X_1, \dots, X_n be a random sample from the probability distribution having density $f(x)$ given by (1). The log-likelihood function $L(\tau)$ is given by

$$L(\tau | X_1, \dots, X_n) = -n\eta(\tau) + n \sum_{i=1}^m \tau_i g_i,$$

where the sample means g_i are defined by

$$g_i = \sum_{j=1}^n \gamma_i(X_j) / n, \quad i = 1, 2, \dots, m.$$

Since $H(\tau)$ is positive definite, it follows (Zangwill, 1969) that $\eta(\tau)$ is strictly convex over R^m . Therefore $L(\tau)$ is strictly convex over R^m , and so the maximum likelihood estimator (when it exists) must be the unique solution of the equation $\Delta L(\tau) = 0$. Now let g be the vector (g_1, \dots, g_m) , and note that

$$\Delta L(\tau) = n[g - \Delta\eta(\tau)], \quad \Delta^2 L(\tau) = -n\Delta^2\eta(\tau) = -nH(\tau).$$

From this it may be seen that the maximum likelihood estimator (MLE) of τ must satisfy the equation $g = \Delta\eta(\tau)$.

In Crain (1976) it is shown that one can determine the minimal sample size necessary to ensure the existence (with probability 1) of the maximum likelihood estimator, given certain regularity conditions on the functions $\{\gamma_i(x)\}$. Wilks (1962) gives regularity conditions under which the maximum likelihood estimator is efficient and asymptotically normal. These regularity conditions are met by the parametric family of densities given in (1). Also, Lehmann (1959) indicates that g is a complete sufficient statistic for τ . Thus maximum likelihood estimation would seem to be the best way to estimate τ , and if the MLE of τ can be computed reasonably then one should endeavor to do so. Unfortunately, there is no general closed form expression for $\eta(\tau)$ other than

$$\eta(\tau) = \log \left\{ \int_a^b \exp \left[\sum_{i=1}^m \tau_i \gamma_i(x) \right] dx \right\}.$$

This means that to compute the MLE of τ even approximately, one must numerically maximize the concave function $L(\tau)$ using some iterative scheme. For example, the Newton-Raphson method generates successive estimates t_1, t_2, \dots for τ by means of the recursion formula

$$t_{n+1} = t_n + [g - \Delta\eta(t_n)][H(t_n)]^{-1}.$$

Notice that within each iteration the moment vector $\Delta\eta(t_n)$ and the covariance matrix $H(t_n)$ must be computed based on the current parameter estimate t_n of τ . However, since closed form expressions for these moments do not in general exist, they must be prohibitively time-consuming to calculate. In this paper we present a direct, *non-iterative* method of estimating the parameter vector τ , given one additional restriction on the functions $\{\gamma_i(x)\}$.

2. ESTIMATION OF τ

In the previous section it was assumed that $\{\gamma_i(x)\}$ are measurable functions bounded on $[a, b]$, such that $\gamma_0(x), \dots, \gamma_m(x)$ are linearly independent, and $\gamma_0(x)$ is identically 1 on $[a, b]$. We now assume further that $\gamma_i \in C^2[a, b]$, that is, we assume that the second derivative $\gamma_i''(x)$ exists and is continuous for $x \in [a, b]$ and $i = 0, \dots, m$.

Taking the logarithmic derivative of the density (1), we obtain for $x \in [a, b]$:

$$d[\log f(x)]/dx = f'(x)/f(x) = \sum_i \tau_i \gamma_i'(x).$$

Let $\beta \in \mathbb{R}^m$, and consider the function $Q(\beta)$ defined by

$$Q(\beta) = \int_a^b [f'(x)/f(x) - \sum_i \beta_i \gamma_i'(x)]^2 f(x) dx. \quad (2)$$

Clearly $Q(\beta) \geq 0$ for all β , and $Q(\tau) = 0$. Therefore $Q(\beta)$ has a global minimum at $\beta = \tau$, which implies that $\Delta Q(\tau) = 0$. For arbitrary β , (2) yields $\Delta Q(\beta) = 2(\beta M - \alpha)$, where

$$\begin{aligned} M_{ij} &= \int_a^b \gamma_i'(x) \gamma_j'(x) f(x) dx, \\ \alpha_i &= \int_a^b [f'(x)/f(x)] \gamma_i'(x) f(x) dx. \end{aligned} \quad (3)$$

Since $\Delta Q(\tau) = 0$, we then have the *exact* relationship $\tau M = \alpha$.

relationship $\tau M = \alpha$ is the key to a direct, non-iterative procedure for estimating τ . If M is invertible and if the elements of M and α can be estimated, then estimators for τ are simply obtained from $\tau = \alpha M^{-1}$. To establish that this can be done, the following lemmas are necessary.

Lemma 1. Let X have the density (1). Then the $m \times m$ matrix M defined by $M_{ij} = E[\gamma_i'(X) \gamma_j'(X)]$ is invertible.

Proof. Let $\delta \in \mathbb{R}^m$. Then, using δ^t to refer to δ -transpose,

$$\int_a^b \left[\sum_{i=1}^m \delta_i \gamma_i'(x) \right]^2 f(x) dx = \delta M \delta^t.$$

Thus $\delta M \delta^t \geq 0$ for all $\delta \in R^m$, and $\delta M \delta^t = 0$ if and only if $\sum_{i=1}^m \delta_i \gamma_i' = 0$. This can happen if and only if, after integrating with respect to x , there exists a constant, say δ_0 , such that $\sum_{i=1}^m \delta_i \gamma_i(x) = \delta_0$. But this is a contradiction of the assumption that $\gamma_0(x), \dots, \gamma_m(x)$ are linearly independent. Therefore $\delta M \delta^t = 0$ if and only if $\delta = 0$, and so M is positive definite, and hence invertible.

Lemma 2. Let X have the density (1). If $\gamma_i \in C^2[a, b]$ then

$$\alpha_i = f(b)\gamma_i'(b) - f(a)\gamma_i'(a) - E[\gamma_i''(X)].$$

Proof. Using integration by parts on (3), we obtain:

$$\begin{aligned} \alpha_i &= \int_a^b [f'(x)/f(x)] \gamma_i'(x) f(x) dx = \int_a^b \gamma_i'(x) f'(x) dx \\ &= \gamma_i'(x) f(x) \Big|_a^b - \int_a^b \gamma_i''(x) f(x) dx, \end{aligned}$$

from which the lemma follows immediately.

Lemma 3. Let X have density (1), and let $c_k = E[\cos\{k\pi(X-a)/(b-a)\}]$. Then

$$\begin{aligned} f(a) &= [1 + 2 \sum_{k=1}^{\infty} c_k] / (b-a), \quad \text{and} \\ f(b) &= [1 + 2 \sum_{k=1}^{\infty} (-1)^k c_k] / (b-a). \end{aligned}$$

Proof. Let $e(x)$ be the even function constructed from $f(x)$ as follows:

$$e(x) = f(a+x) \text{ for } x \in [0, b-a],$$

$$e(x) = f(a-x) \text{ for } x \in [a-b, 0].$$

The Fourier series expansion of $e(x)$ exists and has the form

$$e(x) = a_0/2 + \sum_{k=1}^{\infty} [u_k \cos\{k\pi x/(b-a)\} + v_k \sin\{k\pi x/(b-a)\}]$$

$$\begin{aligned}
 \text{where } u_k &= \int_{a-b}^{b-a} \cos\{k\pi x/(b-a)\} e(x) dx / (b-a) \\
 &= 2 \int_0^{b-a} \cos\{k\pi x/(b-a)\} f(a+x) dx / (b-a) \\
 &= 2 \int_a^b \cos\{k\pi(x-a)/(b-a)\} f(x) dx / (b-a) \\
 &= 2c_k / (b-a) \quad \text{for } k = 0, 1, 2, \dots, \text{ and} \\
 v_k &= \int_{a-b}^{b-a} \sin\{k\pi x/(b-a)\} e(x) / (b-a) \\
 &= 0 \quad \text{for } k = 1, 2, 3, \dots,
 \end{aligned}$$

since sine is an odd function and $e(x)$ is an even function. Using the fact that $c_0 = 1$, the expansion of $e(x)$ becomes

$$e(x) = [1 + 2 \sum_{k=1}^{\infty} c_k \cos\{k\pi x/(b-a)\}] / (b-a).$$

The lemma follows from $f(a) = e(0)$ and $f(b) = e(b-a)$, since $\cos(k\pi) = (-1)^k$.

These lemmas, together with the previously established relationship $\tau M = \alpha$, yield the following theorem.

Theorem. Let X have density (1) with $\gamma_i \in C^2[a, b]$ for $i = 1, \dots, m$. Define

$$c_k = E[\cos\{k\pi(X-a)/(b-a)\}],$$

$$d_a = [1 + 2 \sum_{k=1}^{\infty} c_k] / (b-a),$$

$$d_b = [1 + 2 \sum_{k=1}^{\infty} (-1)^k c_k] / (b-a),$$

$$\alpha_i = d_b \gamma_i'(b) - d_a \gamma_i'(a) - E[\gamma_i''(X)], \quad \text{and}$$

$$M_{ij} = E[\gamma_i'(X) \gamma_j'(X)], \quad \text{for } i = 1, \dots, m, \text{ and } j = 1, \dots, m.$$

Then $\tau = \alpha M^{-1}$.

Proof. Follows immediately from the three lemmas given above.

In practice the constants d_a and d_b are approximated by the finite series

$$d_a \approx [1 + 2 \sum_{k=1}^p c_k] / (b-a), \quad d_b \approx [1 + 2 \sum_{k=1}^p (-1)^k c_k] / (b-a),$$

where $0 < p \leq n$. Larger values of p (up to a maximum of n) will, of course, yield better estimates. The moments c_k for $k = 1, \dots, p$ are estimated by

$$c_k \approx (1/n) \sum_{j=1}^n \cos\{k\pi(X_j - a)/(b-a)\}.$$

Similarly, the matrix M is estimated by

$$M_{ij} \approx (1/n) \sum_{k=1}^n \gamma'_i(X_k) \gamma'_j(X_k),$$

while the vector α is estimated by

$$\alpha_i \approx d_b \gamma'_i(b) - d_a \gamma'_i(a) - (1/n) \sum_{j=1}^n \gamma''_i(X_j).$$

The estimated value of τ so obtained may be used as the first estimate in the Newton-Raphson procedure, if the MLE's are needed.

3. AN EXAMPLE

The generalized Inverse Gaussian probability density function has been used to describe the waiting times between discharges of neurons (Yang and Chen, 1978). In its exponential form it is written

$$f(x) = \exp[\tau_1 \ln(x) + \tau_2 x + \tau_3 x^2 - \eta(\tau)],$$

for $x \in (0, \infty)$. Neurons have a refractory interval after discharge, of length a (say), and a neuron is observed for a finite length of time, say b , after each discharge. Thus the task is to estimate τ given observations of n independent neural discharges, where the density is defined on $[a, b]$. Let $\{X_i\}$

be the observations. In this case the matrix M is:

$$M = \begin{bmatrix} E[X^{-2}] & E[X^{-1}] & 2 \\ E[X^{-1}] & 1 & 2E[X] \\ 2 & 2E[X] & 4E[X^2] \end{bmatrix},$$

and the vector α is

$$\alpha = [-d_b/b + d_a/a + E[X^{-2}], d_b - d_a, 2bd_b - 2ad_a - 2].$$

Note that the estimator αM^{-1} does not require computation of $\eta(\tau)$, nor does it require estimation of any high order moments of X . On the other hand, it does require the calculation of a few sinusoidal moments c_k , in order to obtain approximations for d_a and d_b .

REFERENCES

- Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, New York.
- Crain, B. R. (1976). Exponential models, maximum likelihood estimation, and the Haar condition. *Journal of the American Statistical Association*, 71, 737-740.
- Lehman, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
- Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York.
- Yang, G. L. and Chen, T. C. (1978). On statistical methods of neuronal spike-train analysis. *Mathematical Biosciences*, 38, 1-34.
- Zangwill, W. I. (1969). *Nonlinear Programming: A Unified Approach*. Prentice-Hall, Englewood Cliffs, New Jersey.

[Received May 1980. Revised October 1980]

PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF A MIXING DISTRIBUTION

BRUCE G. LINDSAY

Department of Statistics
The Pennsylvania State University
University Park, Pennsylvania 16802 USA

SUMMARY. Given a random sample from a mixture density, the objectives is to estimate the mixing distribution by maximum likelihood. Certain properties of this estimator are identified. The estimator is characterized by a family of inequalities which correspond to the usual parametric likelihood equations. If the atomic densities underlying the mixture are of exponential type, it is demonstrated that the estimator matches the first sample moment to the first theoretical moment evaluated at the estimator. However, the sample variance is related to the theoretical variance by an inequality. Following this structural analysis, several algorithms for the estimator are discussed. The paper concludes with an example and a discussion of the difficulties of an asymptotic distribution theory.

KEY WORDS. Mixtures, mixing distributions, maximum likelihood estimator, exponential family.

1. INTRODUCTION

Let $f(x; \phi)$: $\phi \in \Omega$ be a family of density functions, with respect to some sigma finite measure, for random variable X . It will be assumed that Ω is a measurable space so that we may define $PM(\Omega)$ to be the family of probability measures on Ω . Typically, Ω will be an interval of the real line with the usual measurable sets. In this setting the function

$$f(x; Q) = \int f(x; \phi) dQ(\phi), \quad Q \text{ in } PM(\Omega),$$

also defines a density for X . If the measure Q were known, it would be the prior; in this setting it is assumed unknown. The family of density functions so generated will be called the *mixture family* generated by density f . The original family of densities are a subfamily of the mixture family generated by using the atom measures $\delta(\phi)$:

$$f(x; \delta(\phi)) = f(x; \phi). \quad (1)$$

Henceforth, it will be assumed that the atom measures are in $PM(\Omega)$. The *atomic densities* (1) correspond to an underlying simple process. Data generated by a non-atomic measure Q corresponds to the observed mixture of simple processes.

The observed data (x_1, \dots, x_n) will be a random sample from the mixture density and the objective will be to estimate Q by finding a probability measure \hat{Q} in $PM(\Omega)$ which maximizes the likelihood:

$$L(Q) = \prod_{i=1}^n f(x_i; Q).$$

If any of the $f(x_i; \phi)$ are unbounded in ϕ , then the likelihood $L(Q)$ is unbounded, and so no maximum exists. Henceforth, it will be assumed that for each i , $f(x_i; \phi)$ is bounded. Notice that it is possible for \hat{Q} to be an atom $\delta(\hat{\phi})$, in which case $\hat{\phi}$ must be the maximum likelihood estimator of ϕ within the atomic family of densities $f(x; \phi)$.

It turns out to be sufficient for maximization purposes to restrict attention to discrete measures Q with a finite set of points of positive probability (Lindsay, 1980). For convenience a dual notation for discrete measures has been adopted herein. Each such measure which has m points of positive probability will be said to have *support size* m . There will be an m -vector of *weights* $\tilde{\pi}$ and an m -vector of *support points* $\tilde{\phi}$ such that $Q = \sum_j \pi_j \delta(\phi_j)$. In this case

$$f(x; Q) = \sum_{j=1}^m \pi_j f(x; \phi_j).$$

A natural issue which arises within the mixture family is the identifiability of the mixture Q . These issues have been explored in Teicher (1963). However, even if the mixtures Q are themselves unidentifiable there generally will be parameters of the mixture system which will be identifiable and estimable by the

method of maximum likelihood. For example, if $f(x;\phi)$ is itself a discrete density for X then the cell probability functions,

$$\int f(x;\phi)dQ(\phi),$$

are estimable in an asymptotically consistent fashion. A particular example of this simple structure is discussed in Section 4.

The major new results of this paper are found in the following section, where the mathematical structure of the maximum likelihood estimator of the mixture is examined. Central to the section is a theorem which characterizes the estimator as satisfying a derivative inequality similar in nature to the likelihood equations of the parametric models. Focus is put upon results holding for members of the exponential class of densities, resulting in first and second moment properties.

The third section of the paper is a review of computational methods for the maximum likelihood estimator. The fourth section is an identification of issues and difficulties associated with establishing a limiting distribution theory to be used for inference in the mixture family. The discussion is motivated by a simple mixture model arising in genetics.

2. STRUCTURAL PROPERTIES

2.1 Prior Results. The first substantial theoretical developments concerning the maximum likelihood estimator of a mixture were by Kiefer and Wolfowitz (1956). The problem, as they stated it, involved an additional parameter, θ , so that the density was of the form

$$\int f(x;\theta,\phi)dQ(\phi).$$

Although their results concerned simultaneous estimation of θ and Q by maximum likelihood, they apply as well to the case where θ is absent. Their paper established quite general conditions for the consistency of the maximum likelihood estimator. Little more was said about its structure, however, and nothing was said about its computation.

Although in the intervening years there was some development of maximum likelihood methods for mixture models of fixed support size m , the next substantial development for the general problem came twenty years later in the work of Simar (1976). Simar exclusively considered the compound Poisson family. For this family it was demonstrated that a maximizing measure \hat{Q} exists and is unique. It is a discrete measure. Simar gives a bound on

support size which has since been improved by Lindsay (1980). Simar also demonstrated consistency: the maximum likelihood estimator \hat{Q} converges weakly with the probability one to the true measure Q .

Laird (1978) extended some of Simar's results to the general problem. In particular, a self-consistency property was identified (to be discussed in Section 2.2) which leads to an extensive set of conditions under which the mixture maximum likelihood estimator was a discrete measure. It was conjectured that for "well-behaved (analytic) unimodal densities" there would be no more support points needed than observations in the sample. Lindsay (1980) has shown this to be true for virtually all densities.

2.2 Properties of Local Maxima. In this section certain properties of local maxima of the likelihood are derived, where the following is meant by local. Fix the size of the support set at m . Now maximize the likelihood over discrete measures of the form:

$$dQ(\phi) = \sum_{j=1}^m \pi_j \delta(\phi_j).$$

A local maximum is defined to be a maximum which is local in the $2m-1$ dimensional parameter space of the parameters (π, ϕ) .

In the following material it is convenient notationally to treat the parameter ϕ as a random variable Φ which takes on the values according to the probability distribution Q . In this setting the mixture density is the marginal distribution of X in the random pair (X, Φ) . The posterior expectation of a function $h(\phi)$ is then:

$$E(h(\Phi) | X=x; Q) = \frac{\int h(\phi) f(x; \phi) dQ(\phi)}{\int f(x; \phi) dQ(\phi)}$$

The following theorem identifies certain consequences of maximization within the fixed support size framework.

Theorem. Let \hat{Q} be an m -point local maximum. (a) (*Self-consistency*) If $g(\phi)$ is an arbitrary function on Ω then

$$n^{-1} \sum_{i=1}^n E\{g(\Phi) | X=x_i; \hat{Q}\} = E(g(\Phi); \hat{Q}). \quad (2)$$

(b) If Ω is an interval of \mathbb{R} and if $U(\phi; x) \equiv \partial(\ln f)/\partial\phi$ is everywhere well-defined, then for any function $h(\phi)$ which is zero on the boundaries of Ω one has

$$n^{-1} \sum_{i=1}^n E\{h(\phi) U(\phi; X) | X=x_i; \hat{Q}\} = 0. \quad (3)$$

Proof. Equations (2) and (3) are the respective likelihood equations for the two parametric families of mixing distributions, each defined for τ in a neighborhood of zero, given by

$$dQ_{\tau}(\phi) \equiv \exp\{\tau g(\phi)\} d\hat{Q}(\phi) / \int \exp\{\tau g(\phi)\} d\hat{Q}(\phi).$$

$$\text{and} \quad dP_{\tau}(\phi) \equiv d\hat{Q}(\phi + \tau h(\phi)).$$

By definition the point $\tau = 0$ must maximize in both families.

Suppose that the atomic density f is of the form

$$f(x; \phi) = \exp[\phi x - c(\phi)] \quad (4)$$

with respect to some sigma finite measure. Then f is a member of the exponential class of distributions, with ϕ the natural parameter. Properties of this family have been treated extensively elsewhere; for example Barndorff-Nielsen (1978).

Corollary (first moment). Suppose f is of the exponential form (4). Any m -point maximum \hat{Q} which has no support points on the boundary of Ω satisfies

$$E\{X; \hat{Q}\} = E\{c'(\phi); \hat{Q}\} = \bar{x}. \quad (5)$$

Proof. Use $h(\phi) = 1$ in (3) and $g(\phi) = c'(\phi)$ in (2) obtain the results.

2.3 Properties of Global Maxima. The special properties of those mixtures in $PM(\Omega)$ which globally maximize the likelihood will now be presented. In this section we consider properties resulting solely from convexity. In the following section regularity conditions will be assumed.

The following characterization theorem can be found, albeit in several pieces, in Simar (1976) for the special case of the Poisson model. For mixture Q in $PM(\Omega)$, define the following function:

$$D(\phi; Q) = \sum_{i=1}^n \{f(x_i; \phi)/f(x_i; Q) - 1\}.$$

The function D is the Gateaux derivative of $\ln L(Q)$ with respect to Q and so the inequality below (6) is the mixture

model version of the usual likelihood equation: $\sum_{i=1}^n U(\hat{\phi}; x_i) = 0$.

Characterization Theorem. (a) The mixture \hat{Q} maximizes the likelihood $L(Q)$ if and only if

$$D(\phi; \hat{Q}) \leq 0 \quad \text{for all } \phi \text{ in } \Omega. \quad (6)$$

(b) The fitted likelihood values $f(x_i; \hat{Q})$ do not depend on the choice of maximizing measure \hat{Q} .

(c) Let $D^*(\phi) = D(\phi; \hat{Q})$, where by part (b) \hat{Q} may be any maximizing measure. Then all maximizing measures have support in the set

$$S = \{\phi: D^*(\phi) = 0\}.$$

Proof. Let P be an arbitrary element of $PM(\Omega)$. Let

$$P_\tau = (1 - \tau)\hat{Q} + \tau P, \quad \tau \in [0, 1].$$

The densities $f(x; P_\tau)$ create a parametric subfamily of the mixture family, with score function:

$$\frac{\partial}{\partial \tau} \ln L(P_\tau) = \sum_{i=1}^n \frac{f(x_i; P) - f(x_i; \hat{Q})}{f(x_i; P_\tau)}$$

and score derivative:

$$\frac{\partial^2}{\partial \tau^2} \ln L(P_\tau) = - \sum_{i=1}^n \frac{(f(x_i; P) - f(x_i; \hat{Q}))^2}{f^2(x_i; P_\tau)}, \quad (7)$$

Part (a). It is clear that if the mixture \hat{Q} is a maximum likelihood estimator, then $\tau = 0$ is a maximal point within all the parametric subfamilies generated by varying P . The concavity of $\ln L(P_\tau)$ in τ , equation (7), ensures that the latter condition is also sufficient for \hat{Q} being maximal. The concavity in τ also implies that the maximum occurs at $\tau = 0$ if and only if

$$0 \geq \frac{\partial}{\partial \tau} \ln L(P_\tau) \Big|_{\tau=0} = \int D(\phi; \hat{Q}) dP(\phi). \quad (8)$$

The last inequality will hold for all P if and only if (6) holds. *Part (b)*. Assume that there exists P such that $L(\hat{Q}) = L(P)$, where \hat{Q} is maximal. The score derivative (7) will be strictly negative unless $f(x_i; \hat{Q}) = f(x_i; P)$ for all i . If it is negative, then strict concavity implies that there exists τ in $(0, 1)$ with more likelihood than the endpoints, a contradiction. *Part (c)*. If ϕ^* is in the support of a maximal measure \hat{Q} , then the parametric subfamily with $P = \delta(\phi^*)$ is defined for τ in an entire neighborhood of zero. Hence, (7) can be strengthened to equality for this P .

Suppose the set $\{x_1, \dots, x_n\}$ consists of K distinct values $\{y_1, \dots, y_k\}$. Let \underline{f} be the transpose of the row vector $(f(y_1; \hat{Q}), \dots, f(y_K; \hat{Q}))$. Since the fitted values \underline{f} are unique, it is clear that if $S = \{\phi_1, \dots, \phi_J\}$ then the maximum likelihood estimator \hat{Q} is unique if and only if there is a unique J -vector of weights $\underline{\pi}$ (elements nonnegative and summing to one) satisfying $A\underline{\pi} = \underline{f}$, where A is the $K \times J$ matrix with (k, j) th entry $f(y_k; \phi_j)$. Thus it is always possible, once the maximization has been done, to determine if the estimator is unique. In particular, if $J \leq K$ and A is of full rank, then \hat{Q} is unique. Lindsay (1980) has shown that there is an *a priori* guarantee of uniqueness for atomic densities of the exponential class.

2.4 Properties Under Regularity. In this section the characterization results of the last subsection are combined with regularity assumptions to obtain several new properties. In this and the following subsection, it will be assumed that Ω is an interval and the atomic densities $f(x; \phi)$ are twice differentiable in ϕ for all x in the sample space. The extensions to vector-valued ϕ are generally clear. Since one can write

$$D^*(\phi) = \sum_{i=1}^n w_i f(x_i; \phi) - n,$$

where the w_i are functions of the data, the regularity of D^* follows from that of f . The following result is a corollary to the characterization theorem.

Corollary. If ϕ^* is in the interior of Ω , then

$$D^*(\phi^*) = 0 \quad \text{and} \quad D^{**}(\phi^*) \leq 0. \quad (9)$$

The characterization theorem can also be used to bound the range of potential support points for Q .

Proposition. Suppose that for each i the function $f(y_i; \phi)$ has a unique mode ϕ_i in Ω . Let ϕ_m be the minimum and ϕ_M be the maximum of $\{\phi_1, \dots, \phi_K\}$. Then \hat{Q} must have its support in the range $[\phi_m, \phi_M]$.

Proof. By the hypothesis $D^*(\phi)$ must be strictly increasing when $\phi < \phi_m$ and strictly decreasing when $\phi > \phi_M$. Since support points are maximal points of D^* , they cannot occur outside the given range.

The generalization of this result to multivariate ϕ is that if each of the functions $f(y_i; \phi)$ has a single mode point ϕ_i , then the support points for Q must lie in the convex hull of the points ϕ_1, \dots, ϕ_K .

2.5 The Second Moment Property. In Section 2.2 it was found that for the exponential class of densities the mixture maximum likelihood estimator, or indeed any local maxima, matched the first sample moment with the theoretical first moment under \hat{Q} . It is now asked if there is a further agreement of moments. After all, the empirical distribution function, which puts mass n^{-1} at each sample point, is both the maximum likelihood estimator and the method of moments estimator of an unknown discrete distribution. The answer for the mixture problem is that there is a second moment inequality rather than an equality. Before turning to the exponential class a general theorem is proved.

Theorem. Let \hat{Q} be a mixing distribution which maximizes the likelihood. If $h(\phi)$ is any function which is zero on the boundary of Ω , or if \hat{Q} has no support points on the boundary of Ω , then

$$n^{-1} \sum_{i=1}^n E\{h^2(\Phi)(U^2(\Phi; X) + U'(\Phi; X)) | X=x_i; \hat{Q}\} \leq 0. \quad (10)$$

Proof. If ϕ_j is the j th element in support of \hat{Q} and it has weight π_j then by inequality (9)

$$0 \geq \pi_j h^2(\phi_j) D^{*''}(\phi_j) = \sum_{i=1}^n \frac{f''(x_i; \phi_j) h^2(\phi_j) \pi_j}{f(x_i; \hat{Q})}.$$

Summing these inequalities over j yields the inequality

$$E\{h^2(\hat{\Phi}) f''(X; \hat{\Phi}) / f(X; \hat{\Phi}) | X=x_1; \hat{Q}\} \leq 0.$$

The proof is finished by noting that $f''(x; \phi) / f(x; \phi) = U^2(\phi; x) + U'(\phi; x)$.

Corollary. (Variance inequality) Let s^2 denote the sample variance. If \hat{Q} has no support points on the boundary of Ω and if the atomic densities are of exponential form (5), then

$$s^2 \leq \text{Var}(X; \hat{Q}). \quad (11)$$

Proof. Use $h(\phi) = 1$ in (10) and then simplify, using equations (2), (3), and (5).

3. COMPUTATION

This section provides a brief introduction to the computational methods currently recommended for the estimation by maximum likelihood of the mixture Q . It will be seen that the mixture is readily estimated, particularly for densities of the exponential class.

3.1 EM Algorithm. This general algorithm was formally identified and characterized by Dempster, Laird, and Rubin (1976), and the application to the fixed support size mixture estimation problem was pointed out there. The algorithm is directly appropriate only for the fixed support size problem. Laird (1978) recommends it, however, based on the observation that since the maximum likelihood estimator is discrete, one need only be sure that one chooses the support size m sufficiently large. It is suggested that one either start with m small and work upwards until the likelihood stops increasing or one start with m large, say near n . The last approach will work because the algorithm converges along a ridge; if m is too large, then it will either drive elements of the π vector to zero or create duplication in the support set.

The technique is iterative and simply described

Step one. Given estimators $(\tilde{\pi}^{(r-1)}, \phi^{(r-1)})$, defining mixture $Q^{(r-1)}$, compute the following conditional probabilities:

$$\pi_{ij}^{(r)} = P\{\phi = \phi_j | X=x_i; Q^{(r-1)}\} = \pi_j^{(r-1)} f(x_i; \phi_j^{(r-1)}) / f(x_i; Q^{(r-1)})$$

for $i = 1, \dots, n$ and $j = 1, \dots, m$.

Step two. The new estimators are

$$\pi_j^{(r)} = \sum_{i=1}^n \pi_{ij}^{(r)} / n, \quad j = 1, \dots, m,$$

and the solutions $\phi_j^{(r)}$ to

$$\sum_{i=1}^n \pi_{ij}^{(r)} U(\phi_j^{(r)}; x_i) = 0, \quad j = 1, \dots, m. \quad (12)$$

If the solutions have not converged, return to step one.

In the exponential class of densities, the score equation (12) gives a weighted mean of the x -sample:

$$c'(\phi_j^{(r)}) = \sum_{i=1}^n \pi_{ij}^{(r)} x_i / \sum_{i=1}^n \pi_{ij}^{(r)}, \quad j = 1, \dots, m.$$

The last equation implies

$$E[c'(\phi); Q^{(r)}] = \sum_{j=1}^m \pi_j^{(r)} c'(\phi_j^{(r)}) = \bar{x},$$

so that the EM algorithm satisfies the moment equation (5) from the first iteration onwards.

The EM algorithm has several attractive features. There are no matrix inversions or complicated calculations to be performed. The likelihood will increase with every iteration, so minima are not stable solutions. Most importantly, as described above, it will converge on a ridge. Its disadvantages include a relatively slow rate of convergence and a potential for terminating at a purely local maximum. There is also the problem of determining a set of starting values (Laird, 1978, recommends an equally spaced grid of ϕ -values) and the wastefulness of searching blindly for the correct support size.

3.2 Stepwise Inclusion. Simar (1976) devised an iterative algorithm for the Poisson model which used the characterization theorem. It could be generalized as follows:

Step one. Given $Q^{(r-1)}$ with support set $S^{(r-1)}$, find a maximum ϕ^* to $D(\phi; Q^{(r-1)})$. If $D(\phi^*; Q^{(r-1)}) \neq 0$, then stop.

Step two. Let $S^{(r)} = S^{(r-1)} \cup \{\phi^*\}$. Maximize the likelihood over the weight vector for the support set $S^{(r)}$, yielding estimator $Q^{(r)}$. Return to step one.

It should be noted that Simar also uses two subalgorithms to reduce the size of the support set if that size exceed certain theoretical bounds. Although this speeds convergence, the maximization over the weight vector in step two will sometimes delete points from $S^{(r)}$ and so the support size is not indefinitely increasing. Lindsay (1980) has improved Simar's bounds and has given bounds for the entire exponential family.

The Simar algorithm has an important advantage; if the estimators converge, then it must be to the full maximum, not a local one. Also, the estimator may be started at a small support size -- for example, the one-point maximum likelihood estimator $\hat{\phi}$ -- and the algorithm will rather quickly adapt if a large number of support points are needed. Although the algorithm monotonely increases the likelihood, Simar noted that the convergence of the algorithm was difficult to prove and did not do so. Other disadvantages include the rather complicated maximization of $D(\phi; Q)$ at each step and the possible inefficiency of adding support points when it is a shift in the support points that is needed: This algorithm, in contrast to the EM algorithm, does not satisfy the first moment equation in one iteration.

3.3 A Synthesis. The computation of the estimator is feasible, but it appears an optimal strategy has not yet been identified. In the interim, a synthesis of the Simar and the Laird approach which combines the computational simplicity of the EM algorithm with the certainty of full, not local, maximization, is as follows: Use the EM algorithm, with a moderate support size m , until convergence to Q' . Then check $D(\phi; Q') \leq 0$. If it is not true, then add to the support of Q' all local maxima ϕ^* of $D(\phi; Q')$ which satisfy $D(\phi^*; Q') > 0$. The logic for this procedure is that for $Q' = \hat{Q}$, the support points are all local maxima. Start with EM algorithm again with the new support set. It should be noted that the expenses of computing the estimator for a large sample size when a large support is needed has yet to be established for this or any other method.

4. LIMITING DISTRIBUTIONS

4.1 Introduction. The method of maximum likelihood has a long history of providing useful and mathematically elegant solutions to problems in estimation. An important part of its usefulness, both in the parametric setting and in the nonparametric setting of the empirical distribution function, is the detailed body of limiting distribution theory to use for statistical inference. The maximum likelihood estimator of a mixture has been shown to be computationally feasible; it is also of mathematical and practical interest because of its blend of the parametric and the nonparametric. However, there is at present no limiting theory to be used for inference. In this section this subject is given a preliminary exploration, motivated by the following example.

4.2 A Trinomial Example. The simplest possible non-trivial mixture model is the trinomial. Let X have the following discrete density:

$$\begin{aligned} f(0; \phi) &= p_0(\phi), \quad f(1; \phi) = p_1(\phi), \\ f(2; \phi) &= p_2(\phi) = 1 - p_0(\phi) - p_1(\phi). \end{aligned}$$

The family of mixture densities is now described by the two parameters

$$p_0(Q) = \int p_0(\phi) dQ(\phi) \quad \text{and} \quad p_1(Q) = \int p_1(\phi) dQ(\phi).$$

The set of permissible cell probabilities under the family of atomic densities is defined by: $G = \{(p_0(\phi), p_1(\phi)) : \phi \in \Omega\}$. The parameter space for the trinomial model is:

$$\Lambda = \{(p_0, p_1) : 0 \leq p_0, \quad 0 \leq p_1, \quad p_0 + p_1 \leq 1\};$$

The parameter space for the mixture trinomial is simply the convex subset of Λ generated by the set G and all convex combinations of points in G . This set is denoted $\text{conv}(G)$. Hence the mixture maximum likelihood estimation of the cell probabilities (p_0, p_1) is simply a constrained version of the full trinomial likelihood problem.

A model of this form can be generated by a simple genetics problem. Suppose that a given locus of a chromosome in a diploid individual can have one of two alleles, A or a . Each individual in the population can then be identified as having one of three genotypes: AA , Aa , or aa . If the population frequency of A is ϕ , and an individual is picked at random from that population,

then according to the Hardy-Weinberg law the probabilities of that individual being AA, Aa, or aa are ϕ^2 , $2\phi(1-\phi)$, and $(1-\phi)^2$ respectively. If we identify $X = 0$ with AA, $X = 1$ with Aa, and $X = 2$ with aa, then a model for a sample from the population is trinomial, with the above cell probabilities. If the sampling is done from a mixture of subpopulations with different gene frequencies ϕ , a state of genetic disequilibrium for the whole population, then the true probabilities of each genotype are mixtures of the Hardy-Weinberg frequencies. The parameters $\int \phi^2 dQ(\phi) \equiv p_0(Q)$ and $\int 2\phi(1-\phi) dQ(\phi) \equiv p_1(Q)$ are identifiable and estimable by maximum likelihood. From them one can estimate the mean and variance of the mixing distribution Q . For the genetics problem the mean of Q is the population frequency and the variance is a measure of disequilibrium.

For the trinomial problem, the maximum likelihood solution proceeds as follows. Identify $\text{conv}(G)$, the convex hull of G . If there are n_0 zeroes in the sample and n_1 ones, then the maximum likelihood estimator of (p_0, p_1) in the full trinomial modal is $(n_0/n, n_1/n)$. It follows that if that point is in $\text{conv}(G)$, then it is also the mixture m.l.e. If it is not in $\text{conv}(G)$ then the concavity of the trinomial log likelihood guarantees that a point on the boundary of $\text{conv}(G)$ maximizes the likelihood. For the genetics problem the boundary consists of G itself, so in this case the mixture m.l.e. has a one-point mixing distribution, $\hat{\phi} = (2n_0 + n_1)/n$. It follows that

$$\hat{p}_0 = \max\{\hat{\phi}^2, n_0/n\}, \quad \hat{p}_1 = \min\{2\hat{\phi}(1-\hat{\phi}), n_1/n\}. \quad (13)$$

4.3 Boundary Difficulties. For parameter values (p_0, p_1) in the interior of Γ , there is no difficulty with the limiting distribution theory for the mixture maximum likelihood estimators (13). For large samples the estimators will almost always be $(n_0/n, n_1/n)$ and the usual trinomial results will take over. However, for atomic mixtures $\delta(\phi)$, whose parameter values are on the boundary of Γ , the theory is different. For example, the statistics $(\hat{\phi}^2, n_0/n)$ will have bivariate normal limiting distribution, and hence \hat{p}_0 , being the maximum of the two, will not be normal. Thus if one does not restrict the parameter space to the interior of Γ one is in the awkward position of having as estimator whose limiting distribution depends on the unknown true distribution in a discontinuous fashion.

This result demonstrates, and results from parametric maximum likelihood theory substantiate, the necessity of identifying an appropriate interior to the mixture parameter space $PM(\Omega)$ before a satisfactory theory can be developed. It is conjectured that distributions of the continuous type should be in that interior. If so, the situation is reminiscent of the theory of the empirical distribution function. There also, the maximum likelihood estimator F_n is discrete, but the distribution-free limit theory for $F_n - F$ depends on F being a continuous distribution.

5. CONCLUDING DISCUSSION

The subject of maximum likelihood estimation of mixtures seems to fall midway between two great bodies of statistical theory: The theory of maximum likelihood in the parametric case and the theory of the empirical distribution function. Each of these subjects has been found to have an elegant and profoundly useful structure. It is anticipated that such a theory could be developed for mixture estimation, which would greatly expand its potential. At present it can only be used in an exploratory fashion. Any further role will depend on the development of an inferential theory.

ACKNOWLEDGEMENTS

This research was partially supported by the National Science Foundation under Grant MCS-8003081.

REFERENCES

- Barndorff-Neilson, O. (1965). Identifiability of mixtures of exponential families. *Journal of Mathematical Analysis and Applications*, 12, 115-121.
- Barndorff-Neilson, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, New York.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39, 1-38.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals of Mathematical Statistics*, 27, 887-906.
- Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73, 805-811.

- Lindsay, B. G. (1980). The geometry of mixture likelihoods. Pennsylvania State University, Department of Statistics Technical Reports and Preprints, 36.
- Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Annals of Statistics*, 4, 1200-1209.
- Teicher, H. (1973). Identifiability of mixtures. *Annals of Mathematical Statistics*, 32, 244-248.

[Received May 1980. Revised October 1980]

CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION IN GAUSSIAN MIXTURES

GEORGE E. POLICELLO II*

Department of Statistics
The Ohio State University

SUMMARY. The paper gives a brief history of the estimation problem for mixtures of gaussian populations. The classical maximum likelihood approach is not appropriate since the likelihood function is unbounded at each sample point. However, this does not seem to cause serious problems when iterative methods are used on a computer. This phenomenon is partially explained by the conditional likelihood approach taken in this paper. In addition, the conditional likelihood approach leads to consistent, asymptotically normal and efficient estimators for the parameters of the mixture. The results of a Monte Carlo study are reported for the univariate case and these show that the procedures provide reasonable estimators for small sample sizes. The methods are then extended to mixtures of multivariate gaussian populations.

KEY WORDS. Gaussian mixtures, maximum likelihood estimation, mixtures of distributions, normal mixtures.

1. INTRODUCTION

In recent years mixture distributions have received considerable attention as models for physically important processes in biology, clinical chemistry, ecology, economics engineering, fisheries, medicine, natural resources, etc. Boswell, Ord, and Patil (1979) provide a thorough discussion of the literature and of some sampling mechanisms leading to mixture distributions. Clinical chemistry is one area where mixture distributions are being studied by principally by nonstatisticians. Gindler (1970) and Grannis and Lott (1978) are two typical examples of the work in this area.

In this paper, the discussion is restricted to mixtures of gaussian populations. Historically, Pearson (1894) appears to be the first person to address the statistical questions associated with a mixture to two univariate gaussian distributions. Unfortunately, Pearson's method of moments approach requires the solution of a ninth order polynomial equation and leads to unstable solutions that could result in negative variance estimates. There were many attempts to overcome these problems. Pearson and Lee (1908-1909) used incomplete moments, Pearson (1915) worked with the moments again, Gottschalk (1948) used half moments, Rao (1948) improved the computational situation slightly and, more importantly, introduced some maximum likelihood alternatives, and Behboodian (1970) continued these efforts. Cohen (1967) created an interesting hybrid approach by introducing a conditional likelihood given the first four moments. Quandt and Ramsey (1978) introduced a clever extension of the method of moments using the moment generating function.

The current resurgence of interest in the problem seems to start with the maximum likelihood estimator (MLE) approach in Hasselbland (1966). Macdonald (1969), Gridgeman (1970) and Fryer and Robertson (1972) are of particular interest. The paper by Day (1969) gives a comprehensive comparison of moment estimators, MLE's and Bayes estimators. Unfortunately, the maximum likelihood procedures usually require an assumption of equal covariance matrices for all of the component populations.

Many of the estimation procedures require iterative methods to obtain the final results. The problems associated with iterative procedures have been studied by Chang (1974), Dick and Bowden (1973), Hasselbland (1966), Hosmer (1973) and Peters and Walker (1978). In a more general setting, Hasselbland (1969) develops iterative procedures for mixtures of distributions in the exponential family.

John (1970) considers the problem of estimating the population of origin for a set of observations on a mixture distribution, hence, relating these procedures to the cluster analysis problems. Rayment (1972) considers the identification problem. James (1978) and Ahmad, Giri and Sinha (1980) restrict their attention to estimation of the mixing parameter. In contrast, Tubbs and Coberly (1976) study the robustness of the mixing proportion estimator when the gaussian components are shifted.

This survey of the literature cannot be completed without mentioning the methods based on graphical techniques. Bhattacharya (1967) provides useful techniques and many references of interest. Other papers dealing with graphical methods are Blischke (1963), Harding (1949), Preston (1953) and Taylor (1965).

In this paper we introduce a conditional maximum likelihood estimator based on ideas different from those in Cohen (1967). For an arbitrary mixture of two gaussian populations, the likelihood function is unbounded at each observation. In the next section we highlight the reasons for the failure of the procedures. Basically, they are trying to estimate the variance of a gaussian population using fewer than two observations. This structure is exploited to create the new estimation procedures in this paper.

2. THE BASIC PROBLEM

The notation required for the multivariate problem is quite complicated and lends nothing to an understanding of the basic principles involved in the estimation procedure. Thus, the theoretical development is restricted to the univariate case and the multivariate results are summarized in a separate section. Also there is no conceptual difficulty in extending the results to mixtures of more than two gaussian distributions; however, the necessary extra steps would unduly lengthen the exposition. Therefore, the main results are obtained for a mixture of two gaussian populations.

Let $\phi(z) = (2\pi)^{-1/2} \exp(-.5z^2)$ denote the standard univariate gaussian density function. For $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and α real numbers such that $\sigma_1^2 > 0, \sigma_2^2 > 0$ and $0 < \alpha < 1$,

$$g(x) = (1-\alpha) \phi[(x-\mu_1)/\sigma_1]/\sigma_1 + \alpha \phi[(x-\mu_2)/\sigma_2]/\sigma_2$$

is the density function for a mixture of two gaussian distributions. Note that the notation suppresses the dependence of $g(x)$ on the parameters. Let the vector of parameters be denoted by

$\theta' = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha)$. The usual demonstration that there is no theoretically reasonable set of MLE's for the five parameters in θ works because it is possible to isolate the term, in the product expansion of the likelihood function, L , that corresponds to one factor from population one (parameters μ_1 and σ_1^2) and $(n-1)$ factors from population two. In this term, there is no companion exponential term involving the variance estimate, $\hat{\sigma}_1^2$, that would attenuate the affect of $\hat{\sigma}_1^2$ going to zero. In effect, including this term in L forces any algorithm to try to estimate the variance of a gaussian population with fewer than two observations. For reasonable sample sizes, situations like this have small probabilities of occurring. Hence, many authors argue that this result is of little practical importance for many of the computer oriented computational schemes; see, especially,

the individual discussions following the Quandt and Ramsey (1978) paper.

The above observations suggest a plan for modifying the MLE procedures to produce an estimation scheme that would overcome the unbounded likelihood problem and retain the, hopefully, good properties of MLE's in general. The four offensive terms above, zero or one observation from one of the populations and all of the remainder from the other population, would not be present if the likelihood were computed conditional on at least two observations from each population. Using this idea, the resulting conditional likelihood function is bounded with probability one. The next section develops the associated estimation procedures.

3. THE CONDITIONAL LIKELIHOOD APPROACH

Take $\theta' = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha)$ and let $x' = (x_1, \dots, x_n)$ be a realization of X , a vector of n independent observations from $g(x)$. It is convenient to introduce a compact notation. Let $\eta = \{1, 2, \dots, n\}$ and $\psi = \{(r_1, r_2, r_3, r_4); (r_1, r_2, r_3, r_4) \text{ is an ordered set of four distinct integers from } \eta\}$. Thus, there are $n!/(n-4)!$ distinct vectors in ψ . Let $\{a_i\}, \{b_i\}, \{c_i\}$ and $\{d_i\}$ be four arbitrary sequences of n real numbers. Then we use $\sum_{\psi} a_i b_j c_k d_\ell$ to denote

$$\sum_{(i,j,k,\ell) \in \psi} a_i b_j c_k d_\ell.$$

Define the event $A = \{\text{at least two } x\text{'s from population one}\}$ and $B = \{\text{at least two } x\text{'s from population two}\}$. Then

$$\Pr\{A \cap B\} = h(\alpha) = 1 - (1-\alpha)^n - \alpha^n - n\alpha(1-\alpha)^{n-1} - n(1-\alpha)\alpha^{n-1}$$

defines $h(\alpha)$. Let $\phi(z)$ be the standard normal pdf and

$$\begin{aligned} t(x; \theta) &= \alpha^2 (1-\alpha)^2 \sum_{\psi} \{\phi[(x_i - \mu_1)/\sigma_1]/\sigma_1\} \{\phi[(x_j - \mu_1)/\sigma_1]/\sigma_1\} \\ &\quad \times \{\phi[(x_k - \mu_2)/\sigma_2]/\sigma_2\} \{\phi[(x_\ell - \mu_2)/\sigma_2]/\sigma_2\} \\ &\quad \times \prod_{m \neq (i,j,k,\ell)} g(x_m) \end{aligned}$$

where $g(x) = (1-\alpha)\phi[(x-\mu_1)/\sigma_1]/\sigma_1 + \alpha\phi[(x-\mu_2)/\sigma_2]/\sigma_2$.

Then the conditional likelihood of θ given $A \cap B$ is proportional to

$$\ell(\theta; x) = t(x; \theta) / [h(\alpha)\alpha^2(1-\alpha)^2].$$

This can be simplified by defining

$$\omega(x) = (1-\alpha)\phi[(x-\mu_1)/\sigma_1]/(\sigma_1 g(x))$$

$$v(x) = \alpha\phi[(x-\mu_2)/\sigma_2]/(\sigma_2 g(x)),$$

and $Q = \prod_{i=1}^n g(x_i).$

Note that for all x , $\omega(x) + v(x) = 1$. Since there is no ambiguity in the argument lists we will use ℓ for $\ell(\theta; x)$, t for $t(x; \theta)$ and h for $h(\alpha)$. The subsequent expressions are notationally simpler with the following additional correspondences, let $\omega_i = \omega(x_i)$ and $v_i = v(x_i)$. Using these new variables

$$t = Q \sum_{\psi} \omega_i \omega_j v_k v_l.$$

In order to maximize ℓ it is necessary to obtain the derivatives of t and h with respect to all of the parameters in θ . For any function of θ , say $p(\theta)$, we denote the various derivatives by

$$p^{\mu_1}_1 = \partial p / \partial \mu_1, \quad p^{\mu_2}_2 = \partial p / \partial \mu_2, \text{ etc.}$$

There is a common structure in all of the ω and v derivatives that can be exploited to reduce the work. Henceforth, let ρ denote an arbitrary element of θ and define

$$q(x; \mu_1) = q(x; \mu_2) = x, \quad q(x; \sigma_1^2) = (x - \mu_1)^2,$$

$$q(x; \sigma_2^2) = (x - \mu_2)^2, \quad \text{and} \quad q(x; \alpha) = v(x).$$

Also define $f(x; \rho) = q(x; \rho) - \rho$, a sign function $s(\rho) = 1$ if $\rho = \mu_1, \sigma_1^2$ or α and $s(\rho) = -1$ if $\rho = \mu_2, \sigma_2^2$

Due to the asymmetry in some of the functions, define

$$\delta(x; \rho) = \omega(x) \quad \text{if} \quad \rho = \mu_1 \quad \text{or} \quad \sigma_1^2, \quad \delta(x; \rho) = v(x) \quad \text{if} \\ \rho = \mu_2 \quad \text{or} \quad \sigma_2^2 \quad \text{and} \quad \delta(x; \alpha) = 1.$$

Finally, define u^0 and η^0 by

$$n u^0 = \sum_{i=1}^n q(x_i; \rho) \delta(x_i; \rho)$$

and $n \eta^0 = \sum_{i=1}^n \delta(x_i; \rho).$

A reasonable amount of calculus yields for all ρ

$$\begin{aligned}
 t^{\rho} &= Q^{\rho} \sum_{\psi} \omega_i \omega_j v_k v_l \\
 &+ 2Q \sum_{\psi} (\omega_i^{\rho} \omega_j v_k v_l + \omega_i \omega_j v_k^{\rho} v_l). \quad (1)
 \end{aligned}$$

For all of the results in this paper, more computational detail is available in Policello (1980).

3.1 *The $\rho \neq \alpha$ Cases.* For $\rho \neq \alpha$ setting $l^{\rho} = 0$ yields

$$\begin{aligned}
 &\rho [n \eta^{\rho} \sum_{\psi} \omega_i \omega_j v_k v_l] + 2 s(\rho) \sum_{\psi} \omega_i v_i \omega_j v_k v_l \\
 &- 2s(\rho) \sum_{\psi} \omega_i v_i \omega_j \omega_k v_l] \\
 &= n u^{\rho} \sum_{\psi} \omega_i \omega_j v_k v_l \\
 &+ 2s(\rho) \sum_{\psi} q(x_i; \rho) \omega_i v_i \omega_j v_k v_l \\
 &- 2s(\rho) \sum_{\psi} q(x_i; \rho) \omega_i v_i \omega_j \omega_k v_l \quad (2)
 \end{aligned}$$

These equations will be used as a part of an iterative scheme for computing the parameter estimates. However, the restricted summations over ψ must be reduced to simple unrestricted summations or the procedure would require too many computation steps to be practical. First define some averages, let

$$n\zeta_{rs} = \sum_{i=1}^n \omega_i^r v_i^s, \quad n\lambda_{rs}^{\rho} = \sum_{i=1}^n q(x_i; \rho) \omega_i^r v_i^s$$

and define γ_{00} by

$$\begin{aligned}
 n^4 \gamma_{00} &= \sum_{\psi} \omega_i \omega_j v_k v_l \\
 &= n^4 \zeta_{10}^2 \zeta_{01}^2 - n^3 [\zeta_{20} \zeta_{01}^2 + \zeta_{10}^2 \zeta_{02} + 4\zeta_{10} \zeta_{11} \zeta_{01}] \\
 &+ n^2 [2\zeta_{21} \zeta_{01} + 2\zeta_{10} \zeta_{12}] - n\zeta_{22}.
 \end{aligned}$$

In practice, it was possible to obtain better numerical results when the r and s subscripts on λ_{rs}^{ρ} and ζ_{rs} were as close as possible. The final results required considerable rearrangement of the initial expressions. To this end, define γ_{rs} by letting $n^{6-(r+s)} \gamma_{rs}$ be the coefficient of λ_{rs}^{ρ} in the following equation:

$$\begin{aligned} & [\sum_{\psi} q(x_i; \rho) \omega_i v_i \omega_j v_k v_l - \sum_{\psi} q(x_i; \rho) \omega_i v_i \omega_j \omega_k v_l] \\ &= \lambda_{11}^{\rho} [n^4 \zeta_{10} \zeta_{01} (\zeta_{01} - \zeta_{10}) + n^3 \zeta_{11} (\zeta_{10} - \zeta_{01}) \\ & \quad + n^2 (\zeta_{12} - \zeta_{21})] \\ & \quad + \lambda_{12}^{\rho} [n^3 \zeta_{10} - 3n^3 \zeta_{10} \zeta_{01} + n^2 \zeta_{10}] \\ & \quad - \lambda_{21}^{\rho} [n^3 \zeta_{01} - 3n^3 \zeta_{10} \zeta_{01} + n^2 \zeta_{01}] \\ & \quad + \lambda_{22}^{\rho} [3n^2 (\zeta_{01} - \zeta_{10})] + \lambda_{32}^{\rho} n - \lambda_{23}^{\rho} n. \end{aligned}$$

For example, $\gamma_{12} = \zeta_{10} - 3\zeta_{10}\zeta_{01} + n^{-1}\zeta_{10}$

and $\gamma_{21} = -[\zeta_{01} - 3\zeta_{10}\zeta_{01} + n^{-1}\zeta_{01}]$. Note that the γ_{rs} are the same for all $\rho \neq \alpha$. Using this notation equation (2) becomes

$$\begin{aligned} & \rho [n\gamma_{00}n^{\rho} \\ & \quad + 2s(\rho)(\gamma_{11}\zeta_{11} + \gamma_{12}\zeta_{12}n^{-1} + \gamma_{21}\zeta_{21}n^{-1} + \gamma_{22}\zeta_{22}n^{-2} \\ & \quad + (\zeta_{32} - \zeta_{23})n^{-3})] \\ &= n\gamma_{00}u^{\rho} \\ & \quad + 2s(\rho)[\gamma_{11}\lambda_{11}^{\rho} + \gamma_{12}\lambda_{12}^{\rho}n^{-1} + \gamma_{21}\lambda_{21}^{\rho}n^{-1} + \gamma_{22}\lambda_{22}^{\rho}n^{-2} \\ & \quad + (\lambda_{32}^{\rho} - \lambda_{23}^{\rho})n^{-3}]. \end{aligned}$$

These equations will be a major part of the iterative computational methods and it will be convenient to have a compact representation for them. Therefore, for $\rho \neq \alpha$, we represent the above equations by $\rho C_{\rho}(x; \theta) = D_{\rho}(x; \theta)$ with the obvious definitions for C_{ρ} and D_{ρ} .

3.2 *The Derivatives when $\rho = \alpha$.* In this case setting $\ell^{\alpha} = 0$ yields

$$\begin{aligned} \alpha &= \zeta_{01} + 2n^{-1}(2\alpha-1) \\ & \quad - (n-1)\alpha^2(1-\alpha)^2[(1-\alpha)^{n-3} - \alpha^{n-3}]/h(\alpha) \\ & \quad - 2n^{-1}\gamma_{00}^{-1}[\gamma_{11}\zeta_{11} + \gamma_{12}\zeta_{12}n^{-1} + \gamma_{21}\zeta_{21}n^{-1} + \gamma_{22}\zeta_{22}n^{-2} \\ & \quad + (\zeta_{32} - \zeta_{23})n^{-3}]. \end{aligned} \tag{4}$$

As in the previous section it is convenient to represent this equation in the form $\alpha C_{\alpha}(x; \theta) = D_{\alpha}(x; \theta)$ but here $C_{\alpha} = 1$.

3.3 The Iterative Computation of $\hat{\theta}$. We propose to calculate the estimate of θ iteratively. Let $\hat{\theta}(k)$ be the estimate computed at the k th step of the procedure. To start the procedure it is necessary to have an initial value, say $\hat{\theta}(0)$. Take $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$. It was found that $\hat{\theta}(0) = (\bar{X} - .1S, \bar{X} + .1S, .25S^2, .25S^2, .5)$ led to satisfactory solutions in a reasonable number of iterations. There may be better choices in some situations but the generally satisfactory performance for this initial vector was sufficient for our purposes. Let $\hat{\theta}'(k) = (\hat{\theta}_1(k), \hat{\theta}_2(k), \hat{\theta}_3(k), \hat{\theta}_4(k), \hat{\theta}_5(k))$. At step $k = 5m + j$, $m = 1, 2, 3, \dots$ and $j = 1, 2, 3, 4, 5$ take

$$\hat{\theta}_{\ell}(5m + j) = \hat{\theta}_{\ell}(5m + j - 1) \quad \text{for } j \neq \ell$$

and

$$\hat{\theta}_j(5m+j) = D_{\theta_j}[x; \hat{\theta}(5m+j-1)]/C_{\theta_j}[x; \hat{\theta}(5m+j-1)] \quad \text{for } j=\ell.$$

This procedure can be viewed as a multivariate modification of the Newton-Raphson method for finding zeros of equations.

4. THE MULTIVARIATE CASE

For $i = 1$ or 2 let ψ_i be the density function for a nonsingular k -variate gaussian distribution with mean vector μ_i and covariance matrix $\Sigma_i = A_i^{-1}$. Let x be an arbitrary real k -vector and $0 < \alpha < 1$, then the multivariate mixture density is given by

$$g(x) = (1-\alpha)\psi_1(x) + \alpha\psi_2(x).$$

Likewise, for the vector x we extend the definitions of ω and v by $\omega(x) = (1-\alpha)\psi_1(x)/g(x)$, and $v(x) = 1 - \omega(x)$.

Continuing this process of extension, let X_1, \dots, X_n denote a random sample of n observation vectors from the distribution with density $g(x)$ and let x_i denote a realization of X_i .

Also define $A = \{\text{at least two } x_i \text{ from population one}\}$,

$B = \{\text{at least two } x_i \text{ from population two}\}$, $Q = \prod_{i=1}^n g(x_i)$,

$\omega_i = \omega(x_i)$, $v_i = v(x_i)$, etc. Then, as before, the conditional

likelihood given $A \cap B$ is proportional to

$$l = Q \sum_{\psi} \omega_i \omega_j v_k v_l / [h(\alpha) \alpha^2 (1-\alpha^2)].$$

Let ρ be the parameter designator, as before, but now ρ takes the values $\mu_1, \mu_2, \Sigma_1, \Sigma_2$ and α . Then the extended definitions of λ_{rs}^{ρ} become

$$n \lambda_{rs}^{\mu_1} = n \lambda_{rs}^{\mu_2} = \sum_{i=1}^n x_i (\omega_i^r) (v_i^s),$$

$$n \lambda_{rs}^{\Sigma_1} = \sum_{i=1}^n (x_i - \mu_1) (x_i - \mu_1)' (\omega_i^r) (v_i^s)$$

$$\text{and } n \lambda_{rs}^{\Sigma_2} = \sum_{i=1}^n (x_i - \mu_2) (x_i - \mu_2)' (\omega_i^r) (v_i^s).$$

Note that $\lambda_{rs}^{\mu_1}$ and $\lambda_{rs}^{\mu_2}$ are k -vectors and $\lambda_{rs}^{\Sigma_1}$ and $\lambda_{rs}^{\Sigma_2}$ are $k \times k$ matrices.

Now using matrix differentiation, some recombination of terms and the direct analogue of the univariate development we arrive at equation (3) and equation (4). Of course, all of the quantities should now be given the appropriate multivariate interpretations. In this fashion the iterative solution procedures in Section 3.3 can be directly applied to the multivariate case.

5. ASYMPTOTIC RESULTS AND RELATED MATTERS

The unconditional likelihood function is given by

$L = Q = \prod_{i=1}^n g(x_i)$. If we formally proceed and try to maximize this by setting the derivatives equal to zero we obtain: for $\rho \neq \alpha$

$$\rho = \sum_{i=1}^n q(x_i; \rho) \delta(x_i; \rho) / \sum_{i=1}^n \delta(x_i; \rho), \quad (5)$$

$$\text{and } \alpha = \tau_{01}, \quad (6)$$

where $q(x_i; \mu_1) = q(x_i; \mu_2) = x_i$, $q(x_i; \Sigma_1) = (x_i - \mu_1)(x_i - \mu_1)'$ and $q(x_i; \Sigma_2) = (x_i - \mu_2)(x_i - \mu_2)'$. The iterative estimates based on equations (5) and (6) are of the EM type (Dempster, Laird and Rubin, 1977) and, hence, converge to a local maximum. In the switching regression setting, Kiefer (1978), has shown that a root of the likelihood equations corresponding to a local maximum is consistent, asymptotically normal and efficient. These

two results, taken together, show that the MLE type procedures will asymptotically produce "best" estimators. In fact, once the solution for equations (5) and (6), say $\tilde{\theta}$, is obtained, it is possible to consistently estimate the covariance matrix for $\tilde{\theta}$ using the inverse of the estimated information matrix (Kiefer, 1978).

Hosmer (1973) has shown that, for "reasonable" sample sizes and initial values, the iterative MLE type estimators will not converge to $\tilde{\theta}$ values associated with the singular points of the likelihood function. Unfortunately, for small sample sizes the singularities can be a problem when the component distributions are not well separated.

Taken together the above observations show that the unconditional MLE type procedures, using equations (5) and (6), lead to consistent, asymptotically normal efficient estimators of θ . We now connect this with the estimators developed in this paper.

Note that equation (3) can be written as

$$\rho + O(n^{-1}) = \sum_{i=1}^n q(x_i; \rho) \delta(x_i; \rho) / \sum_{i=1}^n \delta(x_i; \rho) \quad (7)$$

and equation 4 can be written as

$$\alpha + O(n^{-1}) = \zeta_{01}, \quad (8)$$

where, as usual, $O[g(n)]$ means $\limsup_{n \rightarrow \infty} |O[g(n)]/g(n)| < \infty$.

Thus, for large n , the solution to equations (3) and (4), $\hat{\theta}$, and the MLE type estimator, $\tilde{\theta}$, will differ only by a term of order $O(n^{-1})$. This being the case, $\hat{\theta}$ shares all of the desirable asymptotic properties of $\tilde{\theta}$. Thus, asymptotically $\hat{\theta}$ is consistent, asymptotically normal and efficient.

In equations (7) and (8), the $O(n^{-1})$ terms serve to remove the problems associated with the singularities in the unconditional likelihood approach. Hence, for the MLE type procedures the singularities disappear, as a computational problem, at a rate of $O(n^{-1})$. This, theoretical observation, is consistent with the empirical observations of Wolfe (1965, 1970, 1971), who noticed that the singularities caused computational problems only when $n \alpha \leq 1$.

6. THE MONTE CARLO STUDY

The Monte Carlo study was conducted on the Amdahl 470 computer at the Ohio State University. The SUPER DUPER random number generator (Marsaglia, MacLaren and Bray, 1964) was used to

generate all uniform and gaussian deviates. A random sample of size n from a mixture of a gaussian (μ_1, σ_1^2) distribution and a gaussian (μ_2, σ_2^2) distribution with mixing parameter α ; say X_1, \dots, X_n , was obtained by generating n Uniform $(0,1)$ variates, U_1, \dots, U_n , and n standard gaussian variates; Z_1, \dots, Z_n . Then if $U_i < \alpha$, $X_i = \sigma_2 Z_i + \mu_2$ and if $U_i \geq \alpha$, $X_i = \sigma_1 Z_i + \mu_1$.

The moment generating function approach of Quandt and Ramsey (1978) appears to be the most effective estimation procedure, not based on MLE principles. Hence, we decided to use their parameter selections in order to provide a reasonable basis for comparisons. The seven cases, parameter selections and sample sizes, are given in Table 1. All seven cases in our study used 100 simulation repetitions and the resulting estimated MSE's and estimated standard deviations for the MSE's are reported in Table 2.

In many problems the standard deviations, not the variances, are the quantities of interest. For this reason Table 2 also contains the estimated MSE's for the standard deviations of the populations.

7. SUMMARY

The classical MLE estimation approach cannot be applied to mixtures of gaussian populations since the likelihood function is unbounded at every observation. However, formal application of the method appears to lead to "good" estimators in many cases.

This paper developed a conditional MLE approach given at least two observations from each population in the mixture. Conceptually, some might object to this on the grounds that it is impossible to know when the sample contains at least two observations from each population. Practically, this is no problem. We simply define the estimator to be the solution of the conditional likelihood equations for all situations. It is our contention that this is as reasonable as any other procedure, when there are fewer than two observations from one of the populations and we must estimate all five parameters.

After the procedures were developed they were related to the unconditional, formal, MLE approach. It was shown that the

TABLE 1: *Parameter selection summary.*

Parameter	Case						
	1	2	3	4	5	6	7
μ_1	-3.	-1.	-1.	-3.	-3.	-3.	3.
μ_2	3.	1.	1.	3.	3.	3.	3.
σ_1^2	1.	1.	1.	1.	1.	1.	1.
σ_2^2	3.	3.	3.	16.	36.	36.	.9
α	.5	.5	.5	.5	.25	.75	.5
n	50	50	100	50	50	50	50

TABLE 2: *Estimated mean square errors and average number of iterations to convergence*.*

Parameter	Case						
	1	2	3	4	5	6	7
μ_1	.049 (.011)	.148 (.019)	.095 (.014)	.074 (.013)	.042 (.007)	1.26 (.266)	.331 (.042)
μ_2	.165 (.022)	.553 (.100)	.800 (.140)	1.77 (.298)	5.44 (.794)	7.60 (1.08)	.434 (.054)
σ_1	.032 (.055)	.070 (.016)	.046 (.008)	.061 (.009)	.133 (.101)	3.28 (.536)	.117 (.014)
σ_2	.082 (.009)	.225 (.037)	.240 (.038)	.932 (.134)	2.84 (.492)	2.09 (.310)	.100 (.012)
σ_1^2	.121 (.025)	.316 (.144)	.169 (.024)	.290 (.055)	2.86 (2.73)	94.1 (28.1)	.300 (.029)
σ_2^2	.939 (.115)	1.93 (.245)	1.82 (.224)	40.9 (4.95)	265.5 (33.5)	207.6 (26.6)	.220 (0.18)
α	.006 (.001)	.019 (.003)	0.29 (.004)	.012 (.002)	.012 (.003)	.053 (.006)	.018 (.004)
Iterations	11.1	47.8	69.5	18.9	15.0	41.0	46.5

*Estimated standard deviation for each MSE is given in parentheses below number.

methods only differed by terms of order $O(n^{-1})$. Thus, the conditional approach leads to consistent, asymptotically normal and efficient estimators. The $O(n^{-1})$ term also helps to explain the computational results reported by Wolfe (1965, 1970, 1971).

For the univariate case we report the results of a simulation study showing that the conditional likelihood approach leads to reasonable estimators. It is interesting to note that the conditional likelihood approach is easily modified to cover the switching regressions problem which was also considered in the Quandt and Ramsey paper. However, there does not appear to be an easy modification of the moment generating function approach to cover the multivariate case. Whereas, the conditional likelihood approach leads directly to a solution for the multivariate case.

ACKNOWLEDGEMENTS

The author would like to thank the Instruction and Research Computer Center of the Ohio State University for providing the funds and facilities for the computer work in this paper. This work was partially supported by NIH grant No. 5 R01 GM25452-02.

REFERENCES

- Ahmad, M., Giri, N. and Sinha, B. K. (1980). *Estimation of the mixing proportion of two known distributions*. University of Montreal, Department of Mathematics and Statistics, Report, DMS. No. 80-4.
- Behoodian, J. (1970). On a mixture of normal distributions. *Biometrika*, 57, 215-217.
- Bhattacharya, C. G. (1967). A simple method of resolution of a distribution into gaussian components. *Biometrics*, 23, 115-135.
- Blischke, W. R. (1963). Mixtures of distributions. *Classical and Contagious Discrete Distributions*, G. P. Patil, ed. Statistical Publishing Society, Calcutta. Pages 351-373.
- Boswell, M. T., Ord, J. K. and Patil, G. P. (1979). Chance mechanisms underlying univariate distributions. *Statistical Distributions in Ecological Work*, J. K. Ord, G. P. Patil and C. Taillie, eds. International Co-operative Publishing House, Fairland, Maryland. Pages 3-156.
- Chang, W. C. (1974). Estimation in a mixture of two multivariate normal distributions. *Biometrics*, 30, 377 (Abstract).
- Cohen, A. C. (1967). Estimation in mixtures of two normal distributions. *Technometrics*, 9, 15-28.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, 56, 463-474.

- Dempster, A., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38 (with discussion).
- Dick, N. P. and Bowden, D. C. (1973). Maximum likelihood estimation for mixtures of two normal distributions. *Biometrics*, 29, 781-790.
- Fryer, J. G. and Robertson, C. A. (1972). A comparison of some methods for estimating mixed normal distributions. *Biometrika*, 59, 639-648.
- Gindler, E. M. (1970). Calculation of normal ranges by methods used for resolution of overlapping gaussian distributions. *Clinical Chemistry*, 16, 124-128.
- Gottschalk, V. H. (1948). Symmetric bimodal frequency curves. *Journal of the Franklin Institute*, 245-252.
- Grannis, G. F. and Lott, J. A. (1978). Technique for determining the probability of abnormality. *Clinical Chemistry*, 24, 640-651.
- Gridgeman, N. T. (1970). A comparison of two methods of analysis of mixtures of normal distributions. *Technometrics*, 12, 823-833.
- Harding, J. F. (1949). The use of probability paper for the graphical analysis of polymodal frequency distributions. *Journal of Marine Biological Association U.K.*, 28, 141-153.
- Hasselbland, V. (1969). Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8, 431-444.
- Hasselbland, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64, 1459-1471.
- Hosmer, D. W. Jr. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, 29, 761-770.
- James, I. R. (1978). Estimation of the mixing proportion in a mixture of two normal distributions for simple, rapid measurements. *Biometrics*, 34, 265-275.
- John, S. (1970). On identifying the population of origin of each observation in a mixture of observations from two normal populations. *Technometrics*, 12, 553-563.
- Kiefer, N. M. (1978). Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica*, 46, 427-434.
- Macdonald, P. D. M. (1969). Fortran programs for statistical estimation of distribution mixtures: Some techniques for statistical analysis of length frequency data. Fisheries Research Board of Canada. Technical Report #129.
- Marsaglia, G., MacLaren, M. D. and Bray, T. A. (1964). A fast procedure for generating normal random variables. *Communications of the ACM*, 7, 4-10.

- Pearson, Karl (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society*, 185A, 71-110.
- Pearson, K. (1915). On the problem of sexing osteometric material. *Biometrika*, 40, 479-484.
- Pearson, K. and Lee, A. (1908-1909). On the generalized probable error in multiple normal correlation. *Biometrika*, 6, 59-68.
- Peters, B. C. and Walker, H. F. (1978). Iterative procedure for obtaining maximum likelihood estimates of parameters for a mixture of normal distributions. *SIAM Journal on Applied Mathematics*, 35, 362-378.
- Policello, G. E. II (1980). *A conditional maximum likelihood approach to estimation in gaussian mixtures*. The Ohio State University, Department of Statistics, Technical Report #208, 1-42.
- Preston, E. J. (1953). A graphical method for analysis of statistical distributions into normal components. *Biometrika*, 40, 460-464.
- Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73, 730-752. (With discussion.)
- Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B*, 10, 159-203.
- Rayment, P. R. (1972). The identification problem for a mixture of observations from two normal populations. *Technometrics*, 14, 911-918.
- Robertson, C. A. and Fryer, J. G. (1970). The bias and accuracy of moment estimators. *Biometrika*, 57, 57-65.
- Tan, W. Y. and Chang, W. C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of normal densities. *Journal of the American Statistical Association*, 67, 702-708.
- Taylor, B. J. R. (1965). The analysis of polymodal frequency distributions. *Journal of Animal Ecology*, 34, 445-452.
- Tubbs, J. D. and Coberly, W. A. (1976). An empirical sensitivity study of mixture proportion estimators. *Communications in Statistics - Theory and Methods*, A-5, 1115-1125.
- Wolfe, J. H. (1965). *A computer program for the analysis of tytes*. Technical Bulletin 65-15, U. S. Naval Personnel and Training Research Laboratory, San Diego (Defense Documentation Center AD 620-026).
- Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329-350.
- Wolfe, J. H. (1971). *A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multi-normal distributions*. Technical Bulletin STB 72-2, U. S. Naval Personnel and Training Research Laboratory.

ESTIMATION OF PARETO PARAMETERS BY NUMERICAL METHODS

W. LYLE COOK and DEAN C. MUMME

Department of Mathematics
Idaho State University
Pocatello, Idaho 83209 USA

SUMMARY. For the case when neither of the parameters of a Pareto distribution is known, an iterative process on the unbiased maximum likelihood estimators is investigated. Results from randomly generated samples are compared with other methods of estimation.

KEY WORDS. Pareto distribution, estimation, numerical approximation.

1. INTRODUCTION

The Pareto distribution has as its most celebrated application the modeling of incomes which exceed a minimum value. The cumulative distribution function (c.d.f.) is

$$F(x) = 1 - (a/x)^b, \quad x > a; \quad a, b > 0,$$

and the probability density function is

$$f(x) = ba^b x^{-b-1}.$$

When the parameter a is known, the geometric mean of a sample of size n is sufficient for b , and when b is known the first order statistic is sufficient for a . Together, the first order statistic and the geometric mean form a joint set of sufficient statistics for a and b (Malik, 1970). When either a or b is known, the estimation of the other is relatively straightforward. When neither of the parameters is known, the available estimators may be manipulated so that the value of a

parameter obtained from one can be used in estimating the second. Such a situation suggests a numerical iteration procedure to refine successive estimates. In this paper, we report the results of checking this strategy using artificially generated populations. The results are compared with estimates arrived at through other standard methods.

2. ESTIMATORS

2.1 Least Squares. Let X_1, X_2, \dots, X_n be a random sample from a population following the Pareto distribution and let Y_1, Y_2, \dots, Y_n denote the corresponding order statistics. One estimation involves least squares. By noting that $1 - F(x) = (a/x)^b$ and taking logs we have the expression

$$\log(1 - F(x)) = b \log a - b \log x.$$

This is a linear equation in $y = \log(1 - F(x))$ and $z = \log x$. The standard least squares estimates of the coefficient b and the constant $b \log a$ can be used to approximate the parameters. With pure artificially generated data the least squares estimates return exact values of the parameters. Clearly this limits the suitability of the least squares method in our comparisons.

2.2 Moments. The expected value of a Pareto variable is $ab/(b-1)$ provided $b > 1$. Equating EX with \bar{X} and solving for b yields

$$b = \bar{X}/(\bar{X} - a). \quad (1)$$

Since the c.d.f. of Y_1 is given by

$$F_{Y_1}(x) = 1 - [1 - F_X(x)]^n = 1 - (a/x)^{nb}, \quad x > a,$$

it follows that $EY_1 = anb/(nb-1)$, or

$$a = \frac{(nb-1)Y_1}{nb}. \quad (2)$$

Combining this with (1) gives

$$b = \frac{n\bar{X} - Y_1}{n(\bar{X} - Y_1)}.$$

Thus b can be estimated from Y_1 and \bar{X} , and subsequently used in (2) to estimate a .

2.3 Maximum Likelihood and Sufficient Estimators. Maximizing the likelihood function,

$$L(x_1, \dots, x_n) = \prod_{i=1}^n \frac{b a^b}{x_i^{b+1}},$$

or its logarithm gives an estimate for b of

$$b = \frac{n}{\sum \log X_i - n \log a}.$$

Subject to the constraints on X the likelihood function is maximized in a when $a = Y_1$.

Another method of estimating a and b is to use the joint sufficient pair $(Y_1, \frac{1}{n} \sum \log(X_i/Y_1))$. The equivalent expressions--which would be unbiased if the parameters were fixed--will be referred to as the sufficient estimators. They are

$$a = \frac{(nb - 1)Y_1}{nb}$$

$$b = \frac{n-2}{\sum \log X_i - n \log Y_1}.$$

Here again b is calculated first and then used in the estimate of a . This is a slight variation of the maximum likelihood procedure which estimates a through Y_1 then estimates b through

$$b = \frac{n-1}{\sum \log X_i - n \log a}$$

2.4 Iteration. The original motivation for the investigation was to refine the sufficient estimate by successive iterations from a to b and back, by taking $a_0 = Y_1$ then

$$b_k = \frac{n-1}{\sum \log X_i - n \log a_k}$$

$$a_{k+1} = \frac{(nb_k - 1)Y_1}{nb_k}, \quad k = 0, 1, 2, \dots$$

Each of the values a_1 , b_1 is unbiased when the other parameter is known. Each of the sequences is bounded and monotone and hence converges. It would be desirable to have each converge to the respective unknown parameters, but this would be asking too much of the sample data. A more realistic expectation would be to have the sequences converge to values which improve the initial estimates, and which would be superior to estimates arrived at by other methods.

2.5 Combination. It can be shown that the sequence $\{a_k\}$ is decreasing, thus the estimate arrived at for a through iteration must be less than Y_1 , which is the maximum likelihood estimator. In observing the trial runs, however, it was noticed that quite often this iteration yielded a value less than the known value of a .

An estimator that combined Y_1 with the iterated estimate of a was introduced to partially correct for the occasional undershoot. Since four iterations usually gave an estimate of a which was stable to five decimal places, a ratio of 1:4 was selected. Thus, the combination estimate for a

$$a = \frac{Y_1 + 4(\text{iterated } a)}{5}.$$

3. CRITERIA

Quandt (1966) has performed sampling experiments using moments, maximum likelihood, least squares, and quantiles. His performance criteria were bias, mean square error, and bias in Lorenz coefficients. He found no statistically significant differences based on these criteria, but informally felt the maximum likelihood and quantiles had an edge. He then proposed a method of fitting by minimizing a modified difference function which gave results comparable to the other methods. Our tests compare moments, maximum likelihood, iterations, sufficient, and combination methods. As in Quandt's experiments, bias will be a judgment criteria, but rather than mean square error, we have tabulated the number of times each estimator among the five was closest to the true value of the parameter.

4. SAMPLING

One hundred samples of each size were generated on a Tektronix 4052 by inverting uniformly random data from the unit interval. To account for the random error, which usually exists in real data, the data points were then shifted a scaled random amount left or right according to $y = y + .2(y-a)(\text{RND}[-.5, .5])$. Samples of size 25, 75, 150 were generated and a spectrum of parameter values was used, on the chance that differences might occur as the ratio of a to b changes. Averages were calculated and reported for each method for each set of samples.

5. COMMENTS

The raw sufficient statistic does not compete well for either parameter. This is evident either from a cursory examination of the data or through analysis of the counts, ranks, or averages. There are other statistically significant differences but they do not seem strong enough to adhere to rigidly, especially considering the extreme deviation of the sufficient estimator and the possibility of interaction. Some casual observations and interpretations may be of greater value.

The moments statistic shows very little bias in estimating a , however, it is the closest estimate in relatively few samples. Conversely its bias in estimating b is high, but it is the closest estimator in a good number of samples. The bias readings are consistent with Quandt (1966), but his m.s.e. measurements would not lead a person to expect the closeness readings we observed. The combination statistic, while not a total loss, does not seem to give enough improvement to warrant future consideration. The iteration statistics are not runaway improvements either, but it is safe to say they are competitive, especially when evaluating b . In this case the iteration statistic had the best bias reading and the best closeness count for all values of the parameter selected.

The methods of estimating parameters from a Pareto population utilizing maximum likelihood, iterations, and moments all seem to be of roughly the same order of magnitude of desirability. It may be that maximum likelihood improves with increasing sample size, while the iteration may be of most value when N is small. Moments are equally acceptable and competitive according to our criteria at all sample sizes, and may perform inversely in relation to their bias.

The iteration method certainly is not any worse than standard techniques, and may after more careful analysis be shown to have minor advantages.

TABLE OF PERFORMANCE OF ESTIMATORS
(Numbers in parenthesis record the number of
times the estimator was closest to the true value.)

	Maximum Likelihood	Iterate	Moments	Sufficient	Combination
		a = 1	b = 1.5		
N = 25	$\begin{cases} a & 1.0296 (33) \\ b & 1.5770 (19) \end{cases}$	$\begin{cases} a & 1.0010 (42) \\ b & 1.5104 (43) \end{cases}$	$\begin{cases} a & 1.0053 (7) \\ b & 1.7604 (36) \end{cases}$	$\begin{cases} a & 1.0346 (1) \\ b & 1.7138 (2) \end{cases}$	1.0067 (17)
N = 75	$\begin{cases} a & 1.0098 (32) \\ b & 1.5101 (26) \end{cases}$	$\begin{cases} a & 1.0006 (35) \\ b & 1.4896 (38) \end{cases}$	$\begin{cases} a & 1.0012 (9) \\ b & 1.5984 (30) \end{cases}$	$\begin{cases} a & 0.9989 (2) \\ b & 1.3691 (6) \end{cases}$	1.0024 (22)
N = 150	$\begin{cases} a & 1.0050 (34) \\ b & 1.5229 (24) \end{cases}$	$\begin{cases} a & 1.0005 (28) \\ b & 1.5127 (45) \end{cases}$	$\begin{cases} a & 1.0007 (10) \\ b & 1.5938 (30) \end{cases}$	$\begin{cases} a & 0.9981 (1) \\ b & 1.7272 (1) \end{cases}$	1.0014 (27)
TOTAL	$\begin{cases} a & .0443 (99) \\ b & .1100 (69) \end{cases}$	$\begin{cases} a & .0021 (105) \\ b & .0334 (126) \end{cases}$	$\begin{cases} a & .0081 (26) \\ b & .4526 (96) \end{cases}$	$\begin{cases} a & .0377 (4) \\ b & .5719 (9) \end{cases}$.0106 (66)
BIAS					
		a = 20	b = 5		
N = 25	$\begin{cases} a & 20.1587 (34) \\ b & 5.3951 (17) \end{cases}$	$\begin{cases} a & 19.9960 (34) \\ b & 5.1694 (42) \end{cases}$	$\begin{cases} a & 20.0031 (13) \\ b & 5.3976 (36) \end{cases}$	$\begin{cases} a & 20.1366 (0) \\ b & 5.2803 (5) \end{cases}$	20.0286 (19)
N = 75	$\begin{cases} a & 20.0577 (29) \\ b & 5.1082 (25) \end{cases}$	$\begin{cases} a & 20.0040 (32) \\ b & 5.3901 (40) \end{cases}$	$\begin{cases} a & 20.0048 (13) \\ b & 5.1225 (33) \end{cases}$	$\begin{cases} a & 19.9572 (0) \\ b & 5.2782 (2) \end{cases}$	20.0147 (26)
N = 150	$\begin{cases} a & 20.0250 (35) \\ b & 5.1034 (19) \end{cases}$	$\begin{cases} a & 19.9984 (26) \\ b & 5.0691 (46) \end{cases}$	$\begin{cases} a & 19.9987 (8) \\ b & 5.1304 (33) \end{cases}$	$\begin{cases} a & 19.9893 (0) \\ b & 4.7008 (2) \end{cases}$	20.0038 (31)
TOTAL	$\begin{cases} a & .2414 (98) \\ b & .6066 (61) \end{cases}$	$\begin{cases} a & .0096 (92) \\ b & .2776 (128) \end{cases}$	$\begin{cases} a & .0092 (34) \\ b & .6505 (102) \end{cases}$	$\begin{cases} a & .1901 (0) \\ b & .8578 (9) \end{cases}$.0470 (76)
BIAS					
		a = 6000	b = 2		
N = 25	$\begin{cases} a & 6133.2 (30) \\ b & 2.0965 (16) \end{cases}$	$\begin{cases} a & 6007.0 (33) \\ b & 2.0093 (46) \end{cases}$	$\begin{cases} a & 6017.7 (14) \\ b & 2.1970 (34) \end{cases}$	$\begin{cases} a & 5908.5 (2) \\ b & 1.5961 (4) \end{cases}$	6032.3 (21)
N = 75	$\begin{cases} a & 6043.8 (31) \\ b & 2.0323 (19) \end{cases}$	$\begin{cases} a & 6003.0 (29) \\ b & 2.0048 (43) \end{cases}$	$\begin{cases} a & 6004.1 (11) \\ b & 2.0713 (36) \end{cases}$	$\begin{cases} a & 6026.5 (5) \\ b & 2.2271 (2) \end{cases}$	6011.2 (24)
N = 150	$\begin{cases} a & 6021.4 (38) \\ b & 2.0474 (20) \end{cases}$	$\begin{cases} a & 6001.5 (24) \\ b & 2.0337 (37) \end{cases}$	$\begin{cases} a & 6001.8 (18) \\ b & 2.0684 (41) \end{cases}$	$\begin{cases} a & 6058.0 (2) \\ b & 1.9702 (2) \end{cases}$	6005.5 (18)
TOTAL	$\begin{cases} a & 198.44 (99) \\ b & .1762 (55) \end{cases}$	$\begin{cases} a & 11.53 (86) \\ b & .0467 (126) \end{cases}$	$\begin{cases} a & 23.63 (43) \\ b & .3366 (111) \end{cases}$	$\begin{cases} a & 176 (9) \\ b & .6609 (8) \end{cases}$	48.92 (63)
BIAS					

REFERENCES

- Hastings, N. A. J. and Peacock, J. B. (1975). *Statistical Distributions*. Wiley, New York.
- Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics-Continuous Univariate Distributions - 1*. Wiley, New York.
- Malik, H. J. (1970). Estimation of the parameters of the Pareto distribution. *Metrika*, 16, 126-132.
- Quandt, R. E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika*, 10, 55-82.

[Received June 1980. Revised September 1980]

A NEW ESTIMATION PROCEDURE FOR THE THREE-PARAMETER LOGNORMAL DISTRIBUTION

MARIO CHIEPPA and PANCRAZIO AMATO¹

Instituto di Statistica
and
Istituto di Analisi Matematica
Università di Bari, Italy

SUMMARY. A new procedure for estimating the parameters of the lognormal distribution is presented. A Monte Carlo study, carried out with samples of various sizes, has shown that the procedure is more efficient than maximum-likelihood, when samples are small. The procedure allows, also, hypothesis testing of lognormality. A table of percentage points is presented for this purpose. Finally, an example illustrates the application of the procedure.

KEY WORDS. logarithmic transformation, parameter estimation, test for lognormality.

1. INTRODUCTION

A distribution of X is said to be lognormal if there is a value $\theta \leq X$ such that

$$Z = \log(x-\theta) \sim N(\mu, \sigma^2). \quad (1)$$

The probability density function of a variable X lognormally distributed is

$$p_X = [(X-\theta)\sigma\sqrt{2\pi}]^{-1} \exp\{-[\log(X-\theta)-\mu]^2/2\sigma^2\}. \quad (2)$$

The parameters μ and σ are, respectively, the mean and the standard deviation of the distribution of Z , and θ is a

¹Prof. Amato has contributed with the Appendix.

threshold value below which it is not possible to have values of X .

Many procedures have been proposed to estimate the parameters of a lognormal distribution (Aitchison and Brown, 1957; Johnson and Kotz, 1970; Giesbrecht and Kempthorne, 1976). Among them, the maximum likelihood method estimates the value θ iteratively such that

$$\Sigma(X_i - \hat{\theta})^{-1} + [\Sigma y_i / (X_i - \hat{\theta})] \hat{\sigma} = 0 \quad (3)$$

where:

$$y_i = [\log(X_i - \hat{\theta}) - \hat{\mu}] / \hat{\sigma} \quad (4)$$

$$\hat{\mu} = [\Sigma \log(X_i - \hat{\theta})] / n \quad (5)$$

$$\hat{\sigma} = \{\Sigma [\log(X_i - \hat{\theta}) - \hat{\mu}]^2 / n\}^{1/2} \quad (6)$$

In a previous note (Chieppa and Ricci, 1979), after discussing the meaning and the importance of the parameter θ in the lognormal distribution, a new numerical procedure was suggested for estimating θ . This procedure is now suggested either for the parameters estimation or to test the hypothesis of lognormality.

2. THE PROCEDURE

If a random variable Z is normally distributed, the skewness index is

$$\sqrt{\beta_1} = n^{-1} [\Sigma (Z_i - \mu)^3] / \sigma^3 = 0 \quad (7)$$

and the kurtosis index is

$$\beta_2 = n^{-1} [\Sigma (Z_i - \mu)^4] / \sigma^4 = 3 \quad (8)$$

A value $\hat{\theta}$ can be found by solving

$$\sqrt{\beta_1} = n^{-1} \{\Sigma [\log(X_i - \hat{\theta}) - \hat{\mu}]^3\} / \hat{\sigma}^3 = 0$$

where $\hat{\mu}$ and $\hat{\sigma}$ are given by (5) and (6). Furthermore, if the same value $\hat{\theta}$ satisfies approximately

$$n^{-1} \{\Sigma [\log(X_i - \hat{\theta}) - \hat{\mu}]^4\} / \hat{\sigma}^4 = 3 \quad ,$$

then the distribution of X can be considered reasonably lognormal with $\hat{\theta}$, $\hat{\mu}$ and $\hat{\sigma}$ being the parameter estimates.

Therefore an iterative procedure is suggested for estimating θ , μ and σ , fixing the condition

$$|\sqrt{b_1}| < \varepsilon$$

for some sufficiently small value of ε .

The existence of $\hat{\theta}$, and, consequently, of $\hat{\mu}$ and $\hat{\sigma}$, is demonstrated in the Appendix. If several different θ -estimates exist, then the largest is to be preferred. Afterward, $\hat{\beta}_2$ can be calculated in order to test the hypothesis:

$$H_0: \beta_2 = 3 \quad H_1: \beta_2 \neq 3.$$

3. MONTE CARLO STUDY OF ESTIMATES

In order to check the properties of the suggested procedure, Monte Carlo experiments were performed on the IBM 370/158 computer of Bari University. Random samples of sizes 20, 30, 50, 100, 200, and 500 were generated from a normal population with parameters $\mu = 4$ and $\sigma = 2$. Five hundred samples were generated for each sample size. Each pseudo-random number Z was transformed as follows:

$$X = e^Z + 10, \quad (10)$$

in order to get samples from a lognormal population with parameters $\mu = 4$, $\sigma = 2$ and $\theta = 10$. These parameters have been chosen in order to make useful comparisons with Harter and Moore (1966).

For each sample, parameters θ , μ , σ and β_2 were estimated according to the suggested procedure. For ε the value 0.05 has been fixed; other simulations (Chieppa and Ricci, 1979), find this value convenient. Furthermore, maximum likelihood estimation of the parameters μ , σ , θ and $\sqrt{\beta_1}$ was carried out with the maximum likelihood method, according to Hill (1963).

Means and variances of the estimates, calculated from the 500 samples, using both procedures are given in Table 1.

The results lead to the following conclusions:

a) For the suggested procedure, estimators $\hat{\theta}$, $\hat{\sigma}$ and $\hat{\beta}_2$ show a negative bias, while estimator $\hat{\mu}$ shows a positive bias. Bias and variances decrease as sample size increases; therefore the suggested procedure seems to be consistent. However, the consistency is not a necessary property because of the considerations under b2.

TABLE 1: Means and variances of the estimators computed on 500 random samples, each of size n , generated from a lognormal population with parameters $\mu = 4$, $\sigma = 2$ and $\theta = 10$.

Sample size	suggested procedure		max. likelihood meth.	
	mean	variance	mean	variance
$\hat{\theta}$				
20	8.652	39.5220	-19.981	3292.8442
30	9.241	12.5459	-6.187	3632.2029
50	9.623	3.1648	8.453	770.0303
100	9.950	0.4760	10.346	0.1486
200	9.780	0.2880	10.168	0.0439
500	9.848	0.0735	10.093	0.0116
$\hat{\mu}$				
20	3.971	0.2583	4.249	0.8129
30	4.002	0.1817	4.046	0.4656
50	4.013	0.0859	3.936	0.1202
100	3.994	0.0468	3.952	0.0436
200	4.005	0.0204	3.965	0.0183
500	4.007	0.0094	3.980	0.0087
$\hat{\sigma}$				
20	1.955	0.2147	1.795	0.3220
30	1.964	0.1289	2.010	0.2062
50	1.964	0.0665	2.081	0.0666
100	1.993	0.0257	2.053	0.0235
200	1.963	0.0154	2.023	0.0125
500	1.977	0.0058	2.019	0.0044
$\sqrt{\hat{\beta}_1}$				
20	--	--	0.072	0.9409
30	--	--	-0.146	0.5279
50	--	--	-0.227	0.0780
100	--	--	-0.113	0.0297
200	--	--	-0.086	0.0181
500	--	--	-0.046	0.0076
$\hat{\beta}_2$				
20	2.540	0.4709	3.698	2.9554
30	2.670	0.5093	3.513	2.4640
50	2.712	0.3313	3.098	0.1974
100	2.885	0.2797	3.070	0.0957
200	2.851	0.1703	3.033	0.0529
500	2.868	0.1235	2.986	0.0243

b) A comparison between the suggested procedure and the maximum likelihood method shows that:

b1) When maximum likelihood is used, in the case of small samples, the estimate of θ is more biased and has a larger variance; on the other hand, in the case of large samples, such an estimate is less biased (but with positive bias).

b2) The estimates of μ and σ are substantially the same for both estimation procedures; nevertheless it seems that the suggested procedure is more efficient in the case of small samples; the bias shows an opposite sign.

b3) The estimate of $\sqrt{\beta_1}$ has a negative bias in the case of the maximum likelihood method; no estimate exists, obviously, in the case of the procedure suggested in this paper.

b4) The MLE estimate of β_2 shows a positive bias (the procedure here suggested shows a negative bias); variances, which are larger when sample sizes are small, become smaller when sample sizes increase.

It can be said, in conclusion, that the suggested procedure seems to be more efficient than MLE when samples are small and less efficient when samples are large. The maximum likelihood method yields results in accordance with those obtained by Harter and Moore (1966); the only exception is for $\hat{\theta}$ when n equals 50.

4. TESTING THE HYPOTHESIS OF LOGNORMALITY

In order to test the hypothesis of lognormality, on the basis of the procedure suggested in this paper, a test of the hypothesis:

$$H_0: \beta_2 = 3 \quad H_1: \beta_2 \neq 3$$

is used. The significance limits of β_2 for a normal distribution are well-known (Vianelli, 1959), but they are not useful in this case, as the distribution of β_2 is conditioned by the fixed value of $\sqrt{\beta_1}$ ($\sqrt{\beta_1} < \epsilon$).

Bowman and Shenton (1975) studied the joint distribution of $\sqrt{\beta_1}$ and β_2 in the case of a normal distribution, giving the confidence contours for $\sqrt{\beta_1}$ and β_2 ; but also their significance limits of β_2 , for $\sqrt{\beta_1} = 0$, are not useful in this case.

TABLE 2. Significance limits of β_2 in order to test for lognormality.

Sample size	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
	lower	upper	lower	upper	lower	upper
20	1.68	3.88	1.60	4.27	1.50	4.90
30	1.78	3.85	1.70	4.24	1.60	4.85
40	1.88	3.82	1.79	4.21	1.68	4.80
50	1.97	3.80	1.87	4.18	1.75	4.75
60	2.05	3.78	1.94	4.15	1.81	4.70
70	2.12	3.76	2.00	4.12	1.87	4.66
80	2.18	3.74	2.06	4.10	1.92	4.63
90	2.22	3.72	2.11	4.08	1.97	4.60
100	2.25	3.70	2.15	4.06	2.01	4.57
150	2.28	3.64	2.18	3.96	2.06	4.44
200	2.31	3.60	2.21	3.89	2.10	4.35
300	2.36	3.57	2.27	3.86	2.16	4.31
400	2.41	3.55	2.33	3.83	2.22	4.27
500	2.45	3.53	2.38	3.80	2.28	4.24

From the simulations mentioned in the previous section, percentiles have been calculated in order to obtain the significance limits for $\alpha = 0.10, 0.05$ and 0.01 . These significance limits for different sample sizes are shown in Table 2.

5. AN APPLICATION

In the Istituto di Malattie Infettive of Bari University, DNA-polymerase activity (in counts per minute; c.p.m.) have been observed in 70 chronic carriers of $\text{HB}_{\text{Sg}}\text{A}_{\text{g}}$. The distribution of these values has yielded the following statistics:

$$X_{\min} = 115 \quad X_{\max} = 2044 \quad \bar{X} = 458.1 \quad s = 389.6$$

$$\sqrt{b_1} = 2.30 \quad b_2 = 8.32$$

As $\sqrt{b_1}$ and b_2 are significantly different from 0 and 3, the distribution of DNA-polymerase activity is thought to be log-normal. Therefore, the suggested procedure has been applied for parameter estimation. Results are:

$$\hat{\theta} = 99.1 \quad \hat{\mu} = 5.44 \quad \hat{\sigma} = 0.95$$

$$(\sqrt{\hat{\beta}_1} = -0.02) \quad \hat{\beta}_2 = 3.05$$

As β_2 is not significantly different from 3, the lognormality hypothesis has been accepted and the above values are the parameter estimates. Particularly, $\hat{\theta} = 99$ has the important interpretation of minimal radioactivity (in c.p.m.) existing when analyses for the determination of DNA-polymerase activity were carried out.

Thanks go to M. T. Boswell for assistance with the revision of this manuscript.

REFERENCES

- Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press.
- Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, 62, 243-249.
- Chieppa, M. and Ricci, A. (1979). Un nuovo metodo per la stima del parametro soglia delle distribuzioni lognormali. *Rivista di Statistica Applicata*, 12, 14-19.
- Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. *Journal of the Royal Statistical Society, Series B*, 38, 257-264.
- Harter, H. L. and Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association*, 56, 842-851.
- Hill, B. M. (1963). The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association*, 58, 72-84.
- Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions*, 1. Houghton-Mifflin, Boston.
- Vianelli, S. (1959). *Prontuari per calcoli statistici*. Abbaco, Palermo, 701 and 1362-1363.

[Received July 1980. Revised December 1980]

APPENDIX: EXISTENCE OF θ -ESTIMATES

Lognormal observations tend to be positively skewed while the logarithms of the observations follow a normal distribution with skewness equal to zero. Consider real numbers $x_1 < x_2 < \dots < x_n$ with skewness index

$$\sqrt{b_1}(x) = (\sqrt{n}) [\sum_j (x_j - \bar{x})^3] / [\sum_j (x_j - \bar{x})^2]^{1.5} > 0,$$

and let $z_j(\theta) = \log(x_j - \theta)$, $\theta < x_1$, with skewness index $b(\theta) = \sqrt{b_1}(\theta)$. The object of this Appendix is to demonstrate that there is at least one θ giving a skewness $b(\theta) = 0$.

Theorem. $\lim_{\theta \rightarrow -\infty} b(\theta) = \sqrt{b_1}(x) > 0$, and

$$\lim_{\theta \rightarrow x_1} b(\theta) = -(n-2)/\sqrt{(n-1)} < 0$$

(and therefore there exists at least one θ such that $b(\theta) = 0$).

Proof. Let $Z = \log(X - \theta)$ where the variate X assumes each of the values x_j , $j=1,2,\dots,n$, with probability $1/n$. Recalling that skewness is invariant to linear transformations, and writing

$$Z = \log[1 + (X - x_1)/(x_1 - \theta)] + \log(x_1 - \theta),$$

it is seen that Z has the same skewness as $\log[1 + (X - x_1)/(x_1 - \theta)] \equiv Y$. But as θ approaches $-\infty$, Y becomes indistinguishable from $(X - x_1)/(x_1 - \theta)$ which has the same skewness as X . This proves the first part of the theorem. For the second part, let $M_1 = E[Z^i] = \Sigma z^i/n$. As θ approaches x_1 , the term z_1 approaches $-\infty$, while z_2, z_3, \dots, z_n remain bounded. Hence,

$$M_i/M_1^i \rightarrow n^{i-1}, \quad i=1,2,\dots,$$

and $\text{Var}(Z)/M_1^2 \rightarrow n-1$.

Now

$$b(\theta) = [M_3 - 3M_1 \cdot M_2 + 2M_1^3]/[\text{Var}(Z)]^{1.5}.$$

Dividing numerator and denominator by $M_1^3 = -|M_1|^3$ and using the preceding relations gives

$$b(\theta) \rightarrow -(n^2 - 3n + 2)/(n-1)^{1.5} = -(n-2)/\sqrt{(n-1)}.$$

ON THE ASYMPTOTIC DISTRIBUTION OF THE MULTIVARIATE CRAMER-VON MISES AND HOEFFDING-BLUM-KIEFER-ROSENBLATT INDEPENDENCE CRITERIA

MIKLÓS CSÖRGÖ

Department of Mathematics and Statistics
Carleton University
Ottawa, Canada K1S 5B6

SUMMARY. This exposition consists of two parts. The first one of them is devoted to surveying recent developments in the asymptotic distribution theory of the multivariate Cramér-von Mises statistic. The second part is a preview of some developments in the asymptotic distribution theory of the Hoeffding-Blum-Kiefer-Rosenblatt independence criterion in that the there quoted 1980 joint results with Derek S. Cotterill of the Department of national Defence, Ottawa, have not yet appeared elsewhere.

KEY WORDS. Multivariate Cramér-von Mises statistic, Cramér-von Mises type tests of independence, distribution tables and rates of convergence for both, empirical processes, invariance principles.

1. ON THE LIMITING DISTRIBUTION OF THE MULTIVARIATE CRAMÉR-von MISES STATISTIC

Let Y_1, \dots, Y_n be independent rv uniformly distributed over the d -dimensional unit cube, I^d , ($d \geq 1$), and let $E_n(y)$ be the empirical distribution function of Y_1, \dots, Y_n , i.e., for $y = (y_1, \dots, y_d) \in I^d$, $E_n(y)$ is n^{-1} times the number of $Y_j = (Y_{j1}, \dots, Y_{jd})$, $j = 1, \dots, n$, whose components are less than or equal to the corresponding components of y , conveniently written as

$$E_n(y) = E_n(y_1, \dots, y_d) = n^{-1} \sum_{j=1}^n \prod_{i=1}^d I_{[0, y_i]}(y_{ji}),$$

where I_A denotes the indicator function of the set A .

Consider the uniform empirical process

$$\alpha_n(y) = n^{1/2} [E_n(y) - \lambda(y)], \quad y \in I^d, \quad d \geq 1,$$

where $\lambda(y) = \prod_{i=1}^d y_i$. It will be convenient for us to also think

about $\alpha_n(\cdot)$ in terms of continuous distribution functions F on R^d . Let F be the class of continuous distribution functions on d -dimensional Euclidean space R^d , ($d \geq 1$), and let F_0 be the subclass consisting of every member of F which is a product of its associated one-dimensional marginal distribution functions. Let X_1, \dots, X_n be independent random d -vectors with a common distribution function $F \in F$. Let $F_n(x)$ be the empirical distribution of X_1, \dots, X_n , i.e., for $x = (x_1, \dots, x_d) \in R^d$, $F_n(x)$ is n^{-1} times the number of $X_j = (X_{j1}, \dots, X_{jd})$, $j = 1, \dots, n$, whose components are less than or equal to the corresponding components of x , namely

$$F_n(x) = F_n(x_1, \dots, x_d) = n^{-1} \sum_{j=1}^n \prod_{i=1}^d I_{(-\infty, x_i]}(X_{ji}). \quad (1)$$

Consider the empirical process

$$\beta_n(x) = n^{1/2} (F_n(x) - F(x)), \quad x = (x_1, \dots, x_d) \in R^d, \quad d \geq 1.$$

Let $y_i = F_i^{-1}(x_i)$ be the i th marginal distribution function of $F \in F$ and let $F_i^{-1}(\cdot)$ be its inverse. Now if $F \in F_0$, then

$$\begin{aligned} \beta_n(x) &= n^{1/2} \left[F_n(x) - \prod_{i=1}^d F_i(x_i) \right] \\ &= n^{1/2} \{ F_n[F_1^{-1}(y_1), \dots, F_d^{-1}(y_d)] - \lambda(y) \} \\ &= n^{1/2} [E_n(y) - \lambda(y)] \\ &= \alpha_n(y), \quad y = (y_1, \dots, y_d) \in I^d, \quad d \geq 1, \end{aligned} \quad (2)$$

i.e., if $F \in \mathcal{F}_0$, then β_n is distribution free.

As to $\alpha_n(\cdot)$, the following results are known.

Theorem A. Let X_1, \dots, X_n ($n = 1, 2, \dots$) be independent random d -vectors with a common distribution function $F \in \mathcal{F}_0$ and let $\alpha_n(\cdot)$ be as in (2). Then one can construct a probability space (Ω, \mathcal{A}, P) with $\{\alpha_n(y) : y \in I^d$ ($d \geq 1$), $n = 1, 2, \dots\}$, a sequence of Brownian bridges $\{B_n(y) : y \in I^d$ ($d \geq 1$)} and a Kiefer process $\{K(y, t) : y \in I^d$ ($d \geq 1$), $t \geq 0\}$ on it so that for any $\mu > 0$ there exist a $C > 0$ such that (cf. Csörgő and Révész, 1975) for each n and for $d \geq 1$

$$P\left\{\sup_{y \in I^d} |\alpha_n(y) - B_n(y)| > C(\log n)^{3/2} n^{\frac{-1}{2(d+1)}}\right\} \leq n^{-\mu} \quad (3)$$

and whence

$$\begin{aligned} \sup_{y \in I^d} |\alpha_n(y) - B_n(y)| &\stackrel{a.s.}{=} O\left(n^{\frac{-1}{2(d+1)}} (\log n)^{3/2}\right), \\ \sup_{1 \leq k \leq n} \sup_{y \in I^d} |k^{1/2} \alpha_k(y) - K(y, k)| &\stackrel{a.s.}{=} O\left(n^{\frac{d+1}{2(d+2)}} \log^2 n\right). \end{aligned} \quad (4)$$

Also if $d = 1$, then (c.f. Komlós *et al.*, 1975) for all n and x

$$P\left\{\sup_{0 \leq y \leq 1} |\alpha_n(y) - B_n(y)| > n^{-1/2}(C \log n + x)\right\} > Le^{-\lambda x},$$

where C, L, λ are positive absolute constants [e.g., (cf. Tusnády, 1977a) they can be chosen as $C = 100, L = 10, \lambda = 1/50$],

and

$$P\left\{\sup_{1 \leq k \leq n} \sup_{0 \leq y \leq 1} |k^{1/2} \alpha_k(y) - K(y, k)| > (C \log n + x) \log n\right\} < Le^{-\lambda x},$$

where again C, L, λ are positive absolute constants, and whence

$$\sup_{0 < y < 1} |\alpha_n(y) - B_n(y)| \stackrel{a.s.}{=} O(n^{-1/2} \log n), \quad (5)$$

$$\sup_{1 \leq k \leq n} \sup_{0 \leq y \leq 1} |k^{\frac{1}{2}} \alpha_k(y) - K(y, k)| \stackrel{\text{a.s.}}{=} O(\log^2 n) \quad (6)$$

Further, if $d = 2$, then (c.f. Tusnády, 1977) for all n and x

$$P\left\{\sup_{y \in I^2} |\alpha_n(y) - B_n(y)| > n^{-\frac{1}{2}}(C \log n + x) \log n\right\} < Le^{-\lambda x}, \quad (7)$$

where C, L, λ are positive absolute constants, and whence

$$\sup_{y \in I^2} |\alpha_n(y) - B_n(y)| \stackrel{\text{a.s.}}{=} O(n^{-\frac{1}{2}} \log^2 n).$$

The respective a.s. rates of (3), (4), (6) and (7) are best available, while that of (5) is best possible. For further illuminating comments concerning rates in higher dimensions we refer to Tusnády (1977b).

The Brownian bridges and the Kiefer process of the above theorem are Gaussian processes, defined in terms of a multi-parameter Wiener process as follows:

D1. Wiener process: A separable Gaussian process

$$W(x) = \{W(x_1, \dots, x_d) : 0 \leq x_i < \infty, (i = 1, \dots, d)\}$$

with $EW(x) = 0$ and covariance function

$$EW(x_1)W(x_2) = \lambda(x_1 \wedge x_2),$$

where $x_1 = (x_{11}, x_{12}, \dots, x_{1d})$, $x_2 = (x_{21}, x_{22}, \dots, x_{2d})$,

$$x_1 \wedge x_2 = (x_{11} \wedge x_{21}, \dots, x_{1d} \wedge x_{2d}) \quad \text{and} \quad \lambda(x_1 \wedge x_2) = \prod_{i=1}^d (x_{1i} \wedge x_{2i}).$$

D2. Brownian bridge:

$$B(x) = \{B(x_1, \dots, x_d) : 0 \leq x_i \leq 1 (i = 1, \dots, d)\}$$

$$= \{W(x) - \lambda(x)W(1, \dots, 1) : x \in I^d\} \text{ m with } \lambda(x) = \prod_{i=1}^d x_i.$$

Whence $EB(x) = 0$ and $EB(x_1)B(x_2) = \lambda(x_1 \wedge x_2) - \lambda(x_1)\lambda(x_2)$.

D3. Kiefer process:

$$K(x, t) = \{K(x, t) : x \in I^d, t \geq 0\}$$

$$= \{W(x, t) - \lambda(x)W(1, \dots, 1, t) : x \in I^d, t \geq 0\}.$$

Whence $EK(x, t) = 0$ and

$$EK(x_1, t_1)K(x_2, t_2) = (t_1 \wedge t_2) \{ \lambda(x_1 \wedge x_2) - \lambda(x_1)\lambda(x_2) \}.$$

Given $F \in F_0$, here we are interested in the asymptotic distribution of the multivariate Cramér-von Mises statistic

$$W_{n,d}^2 = \int_{R^d} \beta_n^2(x) \prod_{i=1}^d dF_i(x_i) = \int_{I^d} \alpha_n^2(y) \prod_{i=1}^d dy_i, \quad d \geq 1, \quad (8)$$

where $\beta_n(x)$, $\alpha_n(y)$, $y_i = F_i(x_i)$ are as in (2). Naturally, say by (3), we have for $d \geq 1$ that

$$h(\alpha_n(\cdot)) \xrightarrow{\mathcal{D}} h(B(\cdot)),$$

for every continuous functional h on the space of real valued functions on I^d endowed with the supremum topology, and whence also that

$$W_{n,d}^2 \xrightarrow{\mathcal{D}} W_d^2 = \int_{I^d} B^2(y) dy \stackrel{\mathcal{D}}{=} \int_{I^d} B_n^2(y) dy = W_d^2(n), \quad d \geq 1, \quad (9)$$

with dy standing for $\prod_{i=1}^d dy_i$ from now on. A direct way of seeing (9) is via

$$|W_{n,d}^2 - W_d^2(n)| \stackrel{a.s.}{\rightarrow} \begin{cases} O(r_{1d}(n)(\log \log n)^{\frac{1}{2}}) & \text{if } d \geq 3, \\ O(\rho_2(n)(\log \log n)^{\frac{1}{2}}) & \text{if } d = 2, \\ O(\rho_1(n)(\log \log n)^{\frac{1}{2}}) & \text{if } d = 1, \end{cases} \quad (10)$$

or via

$$|W_{n,d}^2 - n^{-1} \int_{I^d} K^2(y, n) dy| \stackrel{a.s.}{\rightarrow} \begin{cases} O(n^{-\frac{1}{2}} r_{2d}(n)(\log \log n)^{\frac{1}{2}}) & \text{if } d \geq 2 \\ O(\rho_2(n)(\log \log n)^{\frac{1}{2}}) & \text{if } d = 1, \end{cases} \quad (11)$$

where

$$r_{id}(n) = \begin{cases} \frac{1}{n^{2(d+1)}} (\log n)^{3/2} & \text{if } i = 1, \\ \frac{d+1}{n^{2(d+2)}} \log^2 n & \text{if } i = 2, \end{cases}$$

and

$$\rho_i(n) = \begin{cases} n^{-1/2} \log n & \text{if } i = 1 \\ n^{-1/2} \log^2 n & \text{if } i = 2. \end{cases}$$

The respective statements of (10) follow from (3), (7) and (5) respectively and those of (11) by (4) and (6) respectively when they are also combined with appropriate laws of iterated logarithm. From (11), in turn, not only can we deduce that (9) is true, but also a law of iterated logarithm for $W_{n,d}^2$ from that of $\int_{I_d} K^2(y,n)dy$. For a proof of (10) and (11) we refer to that of Corollary 1 in Csörgö (1979).

In addition to (9), from Theorem A we can also prove rates of convergence results for this convergence in distribution. Let $V_{n,d}(x)$ be the distribution function of $W_{n,d}^2$ of (8) and let $V_d(x)$ be that of W_d^2 of (9). Then (9) reads

$$\lim_{n \rightarrow \infty} P\{W_{n,d}^2 \leq x\} = \lim_{n \rightarrow \infty} V_{n,d}(x) = V_d(x), \quad d \geq 1. \quad (12)$$

Put $\Delta_{n,d} = \sup_{0 < x < \infty} |V_{n,d}(x) - V_d(x)|$. Then we have

Theorem B. (Götze, 1979). $\Delta_{n,1} = O(n^{-1+\epsilon})$ for any $\epsilon > 0$.

This theorem is the best available such result for $\Delta_{n,1}$ so far. Earlier S. Csörgö (1976) showed that $\Delta_{n,1} = O(n^{-1/2} \log n)$ and, on the basis of his complete asymptotic expansion for the Laplace transform of $W_{n,1}^2$, he conjectured that $\Delta_{n,1}$ is of order $1/n$. This conjecture was further studied by S. Csörgö and L. Stachó (1980).

As to higher dimensions $d \geq 2$, nothing is known about the exact distribution function $V_{n,d}$ (cf. (12)), and only the characteristic function of V_d (cf. (12)) is known (cf. Dugue, 1969; Durbin, 1970), and that (cf. Anderson and Darling, 1952; Rosenblatt, 1952) W_d^2 may be written in the form

$$W_d^2 = \sum_{k=1}^{\infty} \mu_k^{-1} x_k^2, \quad d \geq 1, \quad (13)$$

where the x_k are independent standard normal random variables and the μ_k are the eigenvalues of the integral equation

$$\int_{I_d} E\{B(x_1)B(x_2)\}f(x_2)dx_2 = \mu f(x_1)$$

with eigenfunctions f and kernel $E\{B(x_1)B(x_2)\}$ (cf. D2.).

Whence, in order to tabulate V_d ($d \geq 2$), one may try working with a numerical inversion of the characteristic function of V_d , or one may try to calculate a number of the necessary eigenvalues for (13). Unfortunately both ways turn out to be quite difficult to follow directly. Durbin (1970) succeeded in solving the latter problem for $d = 2$, and Krivyakova, Martynov and Tyurin (1977) for $d = 3$.

Using the characteristic function of Duge (1969), in Cotterill and Csörgö (1979) we obtain a recursive equation in the cumulants of W_d^2 , and then use the Cornish-Fisher asymptotic expansion to calculate its critical values for $d = 2, 3, \dots, 50$ at various levels of rejection probabilities. These are within 3% of Durbin's values for $d = 2$ and those of Krivyakova, Martynov and Tyurin for $d = 3$. As far as we know there exist no other tables for $d \geq 4$.

Since nothing is known about the exact distribution function $V_{n,d}$ for $d \geq 2$, it is desirable to have a Theorem B type result also for $\Delta_{n,d}$ when $d \geq 2$. As to the latter we have

Theorem C (Cotterill and Csörgö, 1979).

$$\Delta_{n,d} = \begin{cases} O(n^{-1/2} \log^2 n) & \text{if } d = 2, \\ \frac{-1}{O(n^{2(d+1)} (\log n)^{3/2})} & \text{if } d > 2. \end{cases} \quad (14)$$

As far as we know, the rates of (14) are the only available ones for $\Delta_{n,d}$ ($d \geq 2$) and these combined with Theorem B tell the whole story as presently known for $d \geq 1$.

The proof of Theorem C is based on the respective statements of (7) and (3) and on the following two lemmas of Cotterill and Csörgö (1979).

Lemma A. The distribution function V_d of W_d^2 ($d \geq 1$) is arbitrary many times differentiable and for an arbitrary integer p , $V_d^{(p)}(x) \rightarrow 0$ as $x \rightarrow \infty$.

Lemma B. For any real $p \geq 0$ and integers $q = 0, 1, 2, \dots$ the function $x^p V_d^{(q)}(x)$ ($d \geq 1$) is bounded on $(0, \infty)$.

Remark. For $d = 1$, Lemmas A and B were already known (cf. Lemmas 1 and 8 in S. Csörgö (1976)), and we had to prove them only for $d \geq 2$. Also $V_d^{(a)}$ stands for the a th derivative of V_d , and $v_d^{(a)}$ stands for the a th derivative of the density function v_d of W_d^2 ($d \geq 1$).

2. ON THE LIMITING DISTRIBUTION OF THE Hoeffding-Blum-Kiefer-Rosenblatt Independence Criterion

Let again F be the class of continuous distribution functions on d -dimensional Euclidean space R^d , $d \geq 2$, and F_0 be the subclass consisting of every member of F which is a product of its associated one-dimensional marginal distribution functions. Let X_1, \dots, X_n be independent random d -vectors with common unknown distribution $F \in F$, and suppose it is desired to test the null hypothesis

$$H_0: F \in F_0 \text{ against the alternative } H_1: F \in F - F_0. \quad (15)$$

Let again F_n be the empirical distribution function of X_1, \dots, X_n , i.e., for $x = (x_1, \dots, x_d) \in R^d$, $F_n(x)$ is n^{-1} times the number of $X_j = (X_{j1}, \dots, X_{jd})$, $j = 1, \dots, n$, all of whose components are less than or equal to the corresponding components of x , conveniently written as in (1). Let F_{ni} be the marginal empirical distribution function of the i th component of X_j , i.e.,

$$F_{ni}(x_i) = n^{-1} \sum_{j=1}^n I_{(-\infty, x_i]}(X_{ji}), \quad i = 1, \dots, d,$$

and define

$$T_n(x) = T_n(x_1, \dots, x_d) = F_n(x) - \prod_{i=1}^d F_{ni}(x_i), \quad d \geq 2.$$

Hoeffding (1948), when $d = 2$, and Blum, Kiefer and Rosenblatt (1961), when $d \geq 2$, studied the problem of testing H_0 vs. H_1 via appropriate Cramér-von Mises type functionals of their multivariate empirical process $n^{1/2}T_n(x)$, obtaining the characteristic functions of the limiting distributions of these functionals and providing tables for the corresponding distribution functions in the bivariate case. Strong invariance principles for the process $n^{1/2}T_n(x)$ were proved by Csörgö (1979). In order to summarize these asymptotic results and for the sake of stating further ones, we need some further notation.

Let $y_i = F_i^{-1}(x_i)$ be the i th marginal distribution function of F and let $F_i^{-1}(\cdot)$ be its inverse. Define the mapping $L^{-1}: [0,1]^d = I^d \rightarrow R^d$ by

$$L^{-1}(y_1, \dots, y_d) = (F_1^{-1}(y_1), \dots, F_d^{-1}(y_d)), \\ y = (y_1, \dots, y_d) \in I^d \quad (d \geq 2).$$

For our purposes it will be convenient sometimes to view T_n in terms of the latter, and we let

$$t_n(y) = T_n(L^{-1}(y)) = F_n(F_1^{-1}(y_1), \dots, F_d^{-1}(y_d)) - \prod_{i=1}^d F_{ni}(F_i^{-1}(y_i)). \quad (16)$$

Then $H_0: F \in \mathcal{F}_0$ is equivalent to $H_0: F(L^{-1}(y)) = \prod_{i=1}^d y_i = \lambda(y)$, i.e., under H_0 , T_n is distribution free, so we may take F to be the uniform distribution on I^d ($d \geq 2$) and can then study T_n in terms of t_n .

Define the sequence of Gaussian processes $\{T^{(n)}(y): y \in I^d\}$ by

$$\{B_n(y) = \prod_{i=1}^d B_n(1, y_i, 1) \prod_{j \neq i} y_j : y = (y_1, \dots, y_d) \in I^d \ (d \geq 2)\}, \quad (17)$$

where $\{B_n(y) : y \in I^d\}$ is a sequence of Brownian bridges, i.e., for n , B_n is a separable Gaussian process with $EB_n(x) = 0$ and $EB_n(x)B_n(y) = \lambda(x \wedge y) - \lambda(x)\lambda(y)$ ($x, y \in I^d$) (cf. D2).

Define the Gaussian process $\{T(y, t) : y \in I^d, t \geq 0\}$ by

$$\{K(y, t) = \prod_{i=1}^d K(1, y_i, 1, t) \prod_{j \neq i} y_j : y \in I^d \ (d \geq 2), t \geq 0\}, \quad (18)$$

where $\{K(y, t) : y \in I^d, t \geq 0\}$ is a Kiefer process, i.e., a separable Gaussian process with $EK(x, t) = 0$ and $EK(x, s)K(y, t) = (s \wedge t)\{\lambda(x \wedge y) - \lambda(x)\lambda(y)\}$ ($0 \leq s, t < \infty, x, y \in I^d$) (cf. D3).

Obviously $ET^{(n)}(y) = ET(y, t) = 0$, and simple but somewhat tedious calculations yield the covariance functions

$$\begin{aligned} ET^{(n)}(x)T^{(n)}(y) &= \prod_{i=1}^d (x_i \wedge y_i) + (d-1) \prod_{i=1}^d x_i y_i - \sum_{i=1}^d (x_i \wedge y_i) \prod_{j \neq i} x_j y_j \\ &= \rho(x, y) \quad \text{for all } n, \end{aligned} \quad (19)$$

and $ET(x, s)T(y, t) = (s \wedge t)\rho(x, y)$,

where $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d) \in I^d$ and $s, t \geq 0$.

As to $t_n(y) = T_n(L^{-1}(y))$, the following results are known (cf. Theorems 3 and 4 in Csörgö, 1979).

Theorem D. (Csörgö, 1979). Let X_1, \dots, X_n , $n = 1, 2, \dots$, be independent random d -vectors with a common distribution function $F \in \mathcal{F}_0$ and let $t_n(y)$ be as in (16). Then one can construct a probability space (Ω, \mathcal{A}, P) with $\{t_n(y) : y \in I^d \ (d \geq 2), n = 1, 2, \dots\}$, a sequence of Gaussian processes $\{T^{(n)}(y) : y \in I^d \ (d \geq 2)\}$, defined as in (17), and a Gaussian process $\{T(y, t) : y \in I^d \ (d \geq 2), t \geq 0\}$, defined as in (18), on it so that:

(i) If $d = 2$, then for any z and with C, D, μ positive absolute constants

$$P\left\{\sup_{y \in I^2} |n^{\frac{1}{2}} t_n(y) - T^{(n)}(y)| > n^{-\frac{1}{2}}(C \log n + z) \log n\right\} < D e^{-\mu z}, \quad (20)$$

$$\text{whence } \sup_{y \in I^2} |n^{\frac{1}{2}} t_n(y) - T^{(n)}(y)| \stackrel{\text{a.s.}}{=} O(n^{-\frac{1}{2}} \log^2 n).$$

(ii) If $d \geq 3$, then for any $\mu > 0$, there exists a $C > 0$ such that

$$P\left\{\sup_{y \in I^d} |n^{\frac{1}{2}} t_n(y) - T^{(n)}(y)| > C(\log n)^{3/2} n^{-\frac{1}{2}(d+1)^{-1}}\right\} \leq n^{-\mu}, \quad (21)$$

whence

$$\sup_{y \in I^d} |n^{\frac{1}{2}} t_n(y) - T^{(n)}(y)| \stackrel{\text{a.s.}}{=} O(n^{-\frac{1}{2}(d+1)^{-1}} (\log n)^{3/2}), \quad (22)$$

and also

$$\sup_{1 \leq k \leq n} \sup_{y \in I^d} |k t_k(y) - T(y, k)| \stackrel{\text{a.s.}}{=} O(n^{(d+1)/2(d+2)} \log^2 n), \quad d \geq 2. \quad (23)$$

It follows from (17) and (18) that for each n

$$\{T^{(n)}(y): y \in I^d\} \stackrel{D}{=} \{n^{-\frac{1}{2}} T(y, n): y \in I^d\} \stackrel{D}{=} \{T(y, 1): y \in I^d\}. \quad (24)$$

Define the Gaussian process $\{T(y): y \in I^d (d \geq 2)\}$ via (18) by

$$\{T(y): y \in I^d (d \geq 2)\} = \{T(y, t): y \in I^d (d \geq 2), t = 1\},$$

$$\stackrel{D}{=} \{B(y) - \sum_{i=1}^d B(1, y_i, 1) \prod_{j \neq i} y_j: y = (y_1, \dots, y_d) \in I^d (d \geq 2)\} \quad (25)$$

where $\{B(y): y \in I^d\}$ is a Brownian bridge. Thus $T(\cdot)$ has mean zero and the covariance function $\rho(x, y)$ of (19).

By (22), or (23) combined with (24) and (25) we get that under H_0

$$h(n^{\frac{1}{2}}t_n(\cdot)) \xrightarrow{\mathcal{D}} h(T(\cdot))$$

for every continuous functional h on the space of real valued functions on I^d endowed with the supremum topology.

For example, given $F \in F_0$, we get

$$A_{n,d} = n^{\frac{1}{2}} \sup_{y \in I^d} |t_n(y)| \xrightarrow{\mathcal{D}} A_d = \sup_{y \in I^d} |T(y)|, \quad d \geq 2, \quad (26)$$

and

$$\begin{aligned} C_{n,d} &= n \int_{R^d} T_n^2(x) \prod_{i=1}^d dF_1(x_i) \\ &= n \int_{I^d} t_n^2(y) dy \xrightarrow{\mathcal{D}} C_d = \int_{I^d} T^2(y) dy, \quad d \geq 2, \end{aligned} \quad (27)$$

$$\text{with } dy \equiv \prod_{i=1}^d dy_i.$$

As to how near the functionals $A_{n,d}$ and A_d , resp. $C_{n,d}$ and C_d , are to each other, we refer to (4.47), (4.48), resp. to (4.49), (4.50), of Csörgő (1979), which, in turn, also prove (26), resp. (27), directly and they themselves are also a direct consequence of Theorem D.

The calculation of the distribution of the Kolmogorov-Smirnov type rv A_d seems very difficult. Indeed, one does not even know the distribution of $\sup_{y \in I^d} |B(y)|$ nor that of

$$\sup_{y \in I^d} |W(y)|, \quad \text{where } \{B(y): y \in I^d\} = \{W(y) - \lambda(y)W(1): y \in I^d\}$$

is a Brownian bridge and $\{W(y): y \in I^d\}$ is a Wiener process. For sharp inequalities for the distribution of $\sup_{y \in I^d} W(y)$ we refer

to Goodman (1976). It was falsely claimed by Zincenko (1975) that, just like in the case of $d = 1$, we have

$$P\{\sup_{y \in I^d} W(y) > a\} = 2P\{W(1) > a\} \quad \text{also for } d \geq 2. \quad \text{This created}$$

a bit of confusion as to the relevancy of known results (cf. MR52#1912, MR54#11532, 11533).

The calculation of the distribution of the rv C_d is also not easy. For $d = 2$, Blum, Kiefer and Rosenblatt (1961) obtained the characteristic function of C_d and tabulated its

distribution via numerical inversion of its characteristic function. In order to enable ourselves to tabulate approximate critical values of the r.v. $C_{n,d}$ for $d \geq 2$ we (Cotterill and Csörgő, 1980) find the first five cumulants and then use the Cornish-Fisher asymptotic expansion. *Details as to how to calculate approximate critical values for all $d \geq 2$ and tables for the "usual" levels of significance for $d = 2$ to 20 are given in Cotterill and Csörgő (1980).*

Let $\Gamma_{n,d}(\cdot)$ be the distribution function of the rv $C_{n,d}$ and let $\Gamma_d(\cdot)$ be that of the rv C_d . Then (27) reads

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{C_{n,d} \leq x\} &= \lim_{n \rightarrow \infty} \Gamma_{n,d}(x) \\ &= P\{C_d \leq x\} = \Gamma_d(x), \quad d \geq 2. \end{aligned} \quad (28)$$

Since, for various values of n , nothing is known about the exact distribution function $\Gamma_{n,d}$, in addition to (28) it is also of interest to estimate the distance

$$V_{n,d} = \sup_{0 \leq x < \infty} |\Gamma_{n,d}(x) - \Gamma_d(x)|, \quad d \geq 2. \quad (29)$$

The statistic $C_{n,d}$ of (27) cannot be used to test the hypothesis H_0 of (15) unless $F \in F_0$ is completely specified. In some other situations $C_{n,d}$ might come in handy of course. Hoeffding (1948), and Blum, Kiefer and Rosenblatt (1961) suggested, as critical region for H_0 , large values of

$$\hat{C}_{n,d} = n \int_{\mathbb{R}} T_n^2(x) dF_n(x), \quad d \geq 2,$$

or those of

$$\tilde{C}_{n,d} = n \int_{\mathbb{R}} T_n^2(x) \prod_{i=1}^d dF_{n_i}(x_i), \quad d \geq 2. \quad (30)$$

These two statistics are equivalent in that both converge in distribution to the rv C_d of (27).

Recently DeWet (1979) studied a version of (30) in the case of $d = 2$ with some nonnegative weight functions multiplying the integrand T_n^2 of $\tilde{C}_{n,d}$.

Concerning rates of convergence for $\Gamma_{n,d}$ ($d \geq 2$), we have

Theorem E (Cotterill and Csörgö, 1980). For the distance $\nabla_{n,d}$ of (29) we have

$$\nabla_{n,d} = \begin{cases} O(n^{-\frac{1}{2}(d+1)-1} (\log n)^{3/2}) & \text{if } d \geq 3 \\ O(n^{-\frac{1}{2}} \log^2 n) & \text{if } d = 2. \end{cases} \quad (31)$$

As far as we know the rates of (31) for $\nabla_{n,d}$ are the only available ones so far.

The proof of Theorem E is based on the respective statements of (20) and (21) and on the following two lemmas of Cotterill and Csörgö (1980).

Lemma C. The distribution function Γ_d of the rv C_d ($d \geq 2$) is arbitrary many times differentiable and for an arbitrary integer p , $\Gamma_d^{(p)}(x) \rightarrow 0$ as $x \rightarrow \infty$.

Lemma D. For any real $p \geq 0$ and integers $q = 0, 1, 2, \dots$ the function $x^p \gamma_d^{(q)}(x)$ ($d \geq 2$) is bounded on $(0, \infty)$.

Here $\Gamma_d^{(a)}$ stands for the a th derivative of Γ_d and $\gamma_d^{(a)}$ for that of the density function γ_d of C_d ($d \geq 2$).

ACKNOWLEDGEMENTS

This research was supported by a NSERC Canada Grant and a Canada Council Killam Senior Research Fellowship at Carleton University, Ottawa, Canada.

REFERENCES

- Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193-212.
- Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32, 485-498.
- Cotterill, D. S. and Csörgö, M. (1979). *On the limiting distribution of and critical values for the multivariate Cramér-von Mises Statistic*. Carleton Mathematical Lecture Note No. 24.

- Cotterill, D. S. and Csörgő, M. (1980). On the limiting distribution of and critical values for the Hoeffding-Blum-Kiefer-Rosenblatt independence criterion. Manuscript in preparation.
- Csörgő, S. (1976). On an asymptotic expansion for the von Mises ω^2 statistic. *Acta Scientiarum Mathematicarum (Szeged)*, 38, 45-67.
- Csörgő, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process. *Journal of Multivariate Analysis*, 9, 84-100.
- Csörgő, M. and Révész, P. (1975). A new method to prove Strassen type laws of invariance principle II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31, 261-269.
- Csörgő, S. and Stachó, L. (1980). A step toward an asymptotic expansion for the Cramér-von Mises statistic. In *Colloquia Mathematica Societatis János Bolyai*, 21, B. Gyires, ed. Analytic Function Methods in Probability Theory, North-Holland Publishing Company, Amsterdam, Pages 53-65.
- DeWet, T. (1979). Cramér-von Mises tests for independence. Manuscript.
- Dugue, D. (1969). Characteristic functions of random variables connected with Brownian motion and of the von Mises multidimensional ω^2 . In *Multivariate Analysis*, Vol. 2, P. R. Krishnaiah, ed.ⁿ Academic Press, New York. Pages 289-301.
- Durbin, J. (1970). Asymptotic distributions of some statistics based on the bivariate sample distribution function. In *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed. Cambridge University Press. Pages 435-451.
- Götze, F. (1979). *Asymptotic expansions for bivariate von Mises functionals*. Preprints in Statistics 49, University of Cologne.
- Goodman, V. (1976). Distribution estimates for functionals of the two-parameter Wiener process. *Annals of Probability*, 4, 977-982.
- Hoeffding, W. (1949). A nonparametric test of independence. *Annals of Mathematical Statistics*, 19, 546-557.
- Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample D.F.I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 111-131.
- Krivyakova, E. N., Martynov, G. V. and Tyurin Yu. N. (1977). On the distribution of the ω^2 statistic in the multidimensional case. *Theory of Probability and its Applications*, 22, 406-410.
- Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *Annals of Mathematical Statistics*, 23, 617-623.
- Tusnády, G. (1977a). *A study of statistical hypotheses* (in Hungarian). Candidatus Dissertation, Hungarian Academy of Sciences.

- Tusnády, G. (1977b). Strong invariance principles. In *Recent Developments in Statistics*, J. R. Barra et al., eds. North-Holland Publishing Company, Amsterdam. Pages 289-300.
- Tusnády, G. (1977). A remark on the approximation of the sample DF in the multidimensional case. *Periodica Mathematica Hungarica* 8, 53-55.
- Zincenko, N. M. (1975). On the probability of exit of the Wiener random field for some surface (in Russian with English summary). *Teorija Verojatnostei i Matematičeskaja Statistika (Kiev)*, 13, 62-69.

[Received April 1980]

COMPLETE INDEPENDENCE IN THE MULTIVARIATE NORMAL DISTRIBUTION

GOVIND S. MUDHOLKAR

University of Rochester

PERLA SUBBAIAH

Oakland University

SUMMARY. Testing complete independence is one of the simplest problems concerning the covariance structure of a set of measurements. A stepwise procedure proposed by Roy and Bargmann (1958) and a trace criterion due to Nagao (1973) are two well-known competitors of the likelihood ratio test of the hypothesis derived assuming the multivariate normality. We consider some modifications of the Roy-Bargmann procedure based on combinations of independent tests and find them to be asymptotically equivalent to the likelihood ratio test, which is optimal in terms of the exact slopes. The operating characteristics of various tests with samples of moderate size are examined empirically.

KEY WORDS. Combination of tests, exact slopes, stepdown procedure.

1. INTRODUCTION AND SUMMARY

Let X_1, X_2, \dots, X_N be a random sample from a p -variate normal population with covariance matrix Σ . One of the simplest problems concerning the covariance structure of the multivariate normal distribution is of testing the complete independence of

Research supported in part by the Air Force Office of the Scientific Research, Air Force Systems Command, USAF under Grant No. AFOSR-77-3360. The United States Government is authorized to reproduce and distribute reprints for Governmental Purposes notwithstanding any copyright notation hereon.

the p measurements comprising the vectors $\underline{X}_{\sim 1}$. The likelihood ratio test for the complete independence which depends upon the determinant $|\underline{R}_{\sim}|$ of the correlation matrix was derived by Wilks (1935). The exact distribution of the likelihood ratio statistic is discussed and tabulated by Mathai and Katiyar (1979). An alternative solution termed a step-down procedure, which consists of $p-1$ independent tests was proposed by Roy and Bargmann (1958). This procedure, unlike the likelihood ratio test, permits post-hoc analysis of the nature of dependence in case of a rejection of the null hypothesis and depends upon only the well tabulated F-distribution for its implementation.

In this paper we introduce a class of tests asymptotically equivalent, in terms of the exact Bahadur slopes, to the likelihood ratio test which is optimal in this sense. The presently available methods of testing complete independence are summarized in Section 2. The new tests are introduced and shown to be Bahadur-optimal in Section 3. Section 4 contains a Monte Carlo comparison of these tests with the likelihood ratio test and the step-down procedure when the samples are of moderate size. The empirical study also includes a test proposed by Nagao (1973).

2. SOME TESTS OF COMPLETE INDEPENDENCE

Let \underline{R}_{\sim} be the correlation matrix of a sample of size N from the $N_p(\underline{\mu}_{\sim}, \underline{\Sigma}_{\sim})$ population. The likelihood ratio test for

$H_0: \underline{\Sigma}_{\sim} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) = \underline{\Sigma}_d$ rejects it if

$$\Lambda = |\underline{R}_{\sim}|^{-N/2} < c,$$

where the critical constant c may be obtained from Mathai and Katiyar (1979), or obtained by using approximations such as Box's, and Bartlett's discussed by Mudholkar, Trivedi, and Lin (1980). Nagao (1973) noted that asymptotically $-2 \log \Lambda$ is a

χ^2 -variable when the null hypothesis is true but after suitable normalizing it is an asymptotic Gaussian variable for any fixed alternative. He suggested regarding $\tau^2 = \text{tr}(\underline{\Sigma}_{\sim}^{-1} - \underline{I})^2$ which is proportional to the variance of this normal distribution as a noncentrality parameter, i.e., a measure of departure from the null hypothesis, and proposed a consistent estimator

$$T = \frac{N-1}{2} \text{tr}(\underline{S}_{\sim} \underline{S}_d^{-1} - \underline{I})^2 = \frac{N-1}{2} \text{tr}(\underline{R}_{\sim}^2 - \underline{I})$$

of a multiple of τ^2 as a test statistic for H_0 . He obtained the asymptotic expansion for T in the form

$$\Pr(T \leq x) = P_f + \frac{1}{N-1} [a_6 P_{f+6} + a_f P_{f+4} + a_2 P_{f+2} + a_0 P_f] + O(n^{-2}), \quad (1)$$

where $f = p(p-1)/2$ and $P_f = P[\chi_f^2 \leq x]$ and

$a_6 = (p^3 - 3p^2 + 2p)/12$, $a_4 = (-2p^3 + 3p^2 - p)/8$, and $a_2 = (p^3 - p)/4$. He showed that it is satisfactory for $n = 100$.

Roy and Bargmann (1958) consider the null hypothesis of complete independence in the form

$H_0 = \bigcap_{i=2}^p \{H_{oi} : \rho_{i.12...(i-1)}^2 = 0\}$, where $\rho_{i.12...(i-1)}^2$ is the multiple correlation between X_i and $(X_1, X_2, \dots, X_{i-1})$, and note that the sample step-down multiple correlation coefficients $R_{i.12...(i-1)}^2$ are independently distributed when Σ is diagonal.

They propose rejecting the null hypothesis when at least one of the component hypotheses H_{oi} is rejected by the usual test for it, i.e., when $R_{i.12...(i-1)}^2 > a$ constant. This procedure is simple to implement as independently distributed

$$F_i = \{(N-i)/(i-1)\} R_{i.12...(i-1)}^2 / \{1 - R_{i.12...(i-1)}^2\}$$

have variance ratio distributions with $(i-1, N-i)$ d.f., $i = 2, 3, \dots, p$. However, the procedure does require an a priori ordering among the measured variables and a decision regarding the levels α_i of the component tests which, because of the independence, are related to allowable overall type I error α

by $(1-\alpha) = \prod_{i=2}^p (1-\alpha_i)$. It is common to take $\alpha_i = 1 - (1-\alpha)^{1/(p-1)}$

$i = 2, \dots, p$. Roy and Bargmann gave the confidence bounds associated with this step-down procedure which can be used to gain an understanding of the nature of dependence in case H_0 is rejected.

3. A CLASS OF B-OPTIMAL TESTS

The problem of the allocation of the overall type I error among the components tests of the stepdown procedure may be

avoided by considering, instead of the variance ratio statistics F_{i+1} , the P-values P_i associated with the individual tests $i = 1, \dots, k$, where $k = p-1$. Since the statistics F_{i+1} are independent under H_0 , the P-values P_i have independent uniform null distributions. These can therefore be combined variously to construct an overall test for H_0 . The problem of combining independent tests of significance is classical and the literature on the subject is extensive. It is well reviewed in Liptak (1958), Oosterhoff (1969), George (1977) and Mudholkar and George (1979).

A combination procedure for the P-values P_1, P_2, \dots, P_k associated with k independent tests of significance for hypotheses $H_{0i} : \theta_i = \theta_{0i}$ vs. $H_{1i} : \theta_i > \theta_{0i}$, $i = 1, 2, \dots, k$ is based upon a combination statistic $\Psi(P_1, \dots, P_k)$ which is used for testing the overall hypothesis $H_0 = \cap H_{0i}$ vs. the alternative $H_1 = \cup H_{1i}$. The overall null hypothesis H_0 is rejected when $\Psi(P_1, \dots, P_k)$ is large. The following are some of the well known combination statistics: (i) The earliest proposed $\Psi_T = \min \{-2 \log P_i\}$ due to Tippett; (ii) $\Psi_F = \sum -2 \log P_i$ due to Fisher; (iii) $\Psi_N = \sum_i \Phi^{-1}(1-P_i)$, Φ being the c.d.f. of standard normal, considered by Liptak (1958); and (iv) $\Psi_L = \sum_i \log [P_i/(1-P_i)]$ introduced by George (1977). These statistics have simple null distributions. Ψ_T is distributed as the smallest order statistic of a sample from the exponential population, Ψ_F is a χ^2_{2k} -variable, Ψ_N has $N(0, k)$ distribution, and Ψ_L a k -fold convolution of logistic distribution is approximately a t variable with $5k + 4$ degrees of freedom. It is easily seen that the stepdown procedure with equal α_i 's is equivalent to the Tippett combination of its P-values. An account of various studies of the operating characteristics of combination methods in the Neyman-Pearson and decision theoretic framework may be found in Oosterhoff (1969). However, none of the methods can be preferred on the basis of these works. Littel and Folks (1971) examined Bahadur ARE's of various methods and found that among all monotone combination procedures Fisher's is optimal according to this criterion. Mudholkar and George (1979) showed that Ψ_L has the same exact slope as Ψ_F and is consequently optimal. For a recent account of this aspect see

Berk and Cohen (1979). These studies of asymptotic relative efficiencies concern combinations of independent tests; but can be extended to the methods of combining tests which are independent under the null hypothesis only.

The exact slope used in defining Bahadur ARE of a test at an alternative is the rate at which $-2 \log$ (P-value of the test) increases with respect to the sample size n , when the alternative is true. Specifically, let large values of a statistic T_n be significant in testing $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \notin \Theta_0$,

$F_{\theta,n}(t)$ denote the distribution function of T_n and

$P_n(T_n) = 1 - F_{0,n}(T_n)$ be the associated P-value. Then

$c(\theta) = \lim_{n \rightarrow \infty} \{-2n^{-1} \log P_n(T_n)\}$, when it exists, is the exact

slope of T_n . $c(\theta)$ is often obtained using the following result due to Bahadur (1971, p. 27).

Proposition. Suppose that $\lim T_n / \sqrt{n} = b(\theta)$ a.s. for each $\theta \in \Theta - \Theta_0$. Let $\rho_n(t) = -n^{-1} \log[1 - F_{0,n}(\sqrt{n} t)]$ and suppose that

$\lim_{n \rightarrow \infty} \rho_n(t) = \rho(t)$ exists and is continuous on an open interval

containing the range of $b(\theta)$. Then the exact slope of T_n is $c(\theta) = 2 \cdot \rho(b(\theta))$.

Remark. $\rho(t)$ is sometimes referred as the index of the sequence $\{T_n\}$ or of the sequence of distributions $\{F_{0,n}\}$ of $\{T_n\}$.

Now consider the present problem of testing the null hypothesis H_0 that the covariance matrix Σ is diagonal. The step-down procedure which involves testing the component hypothesis $H_{0i}: \rho_{i+1 \cdot 12 \dots i}^2 = 0$ with

$$F_{i+1} = [(N-i-1)/i] \cdot [R_{i+1 \cdot 12 \dots i}^2 / (1 - R_{i+1 \cdot 12 \dots i}^2)], \quad (2)$$

$i = 1, 2, \dots, k$, may be modified by combining these independent (only under H_0) tests using a combination statistic

$\Psi(P_1, \dots, P_k)$, where P_i 's are the P-values associated with

F_{i+1} 's. We are interested in the statistics of the form

$\Psi(P_1, \dots, P_k) = \sum_i \phi_i(P_i)$, where $\phi_i(P_i) = \phi_i(P_{i,n})$ are monotone

decreasing with index $\rho(t) = t$, $i = 1, 2, \dots, k$. Let l denote

the family of these tests. Note that if $\phi_i = G_i^{-1}(1-P_i)$ then under H_0 , $\psi(P_1, \dots, P_k)$ is distributed as the convolution of G_1, G_2, \dots, G_k . In fact L includes in this manner the combinations of the step-down tests based on Fisher's method with $G^{-1}(1-t) = -2 \log t$ and on logit method with $G^{-1}(t) = \log[t/(1-t)]$.

The tests in the family L are asymptotically optimal and equivalent to the likelihood ratio test for H_0 . In order to demonstrate this, i.e., to obtain the exact slope of $\psi(P_1, \dots, P_k) = \Sigma \phi(P_i)$, we examine $\phi_i(P_i)$ which has the same slope as F_{i+1} given in (2). Since F_{i+1} is a variance-ratio with $(i, N-i-1)$ d.f., for an alternative $\rho_{i+1 \cdot 12 \dots i}^2$, $\{F_{i+1}/(N-i-1)\}^{1/2}$ converges in probability to $\{[\rho_{i+1 \cdot 12 \dots i}^2/(1-\rho_{i+1 \cdot 12 \dots i}^2)]/i\}^{1/2}$. This may be rigorously proved using the fact that $i/(n-i+1) F_{i+1}$ is equivalent in law to $[\chi_1^2 + \{Z + \rho_{i+1 \cdot 12 \dots i}^2(1-\rho_{i+1 \cdot 12 \dots i}^2)^{-1/2}\} \chi_{N-i-1}^2]/\chi_{N-i+1}^2$, where $Z, N(0,1)$, and the chi-square variables are mutually independent. Moreover, it can be shown (e.g., see Bahadur, 1971, p. 13) that $-n^{-1} \log(1-F_{0,n}(\sqrt{nt})) \rightarrow \log(1 + i t^2)$. Hence by the above proposition, the exact slope of F_{i+1} or its monotone function $\phi_i(P_i)$ is

$$c_i(\rho_{i+1 \cdot 12 \dots i}^2) = -\log(1-\rho_{i+1 \cdot 12 \dots i}^2).$$

Now, in view of the results by Berk and Cohen (1979) it follows that the index of $\Sigma \phi_i(P_i)$ is the same as the index $\rho_1(t) = t$ of each $\phi_i(P_i)$ and consequently the exact slope of $\psi(P_1, \dots, P_k)$ is

$$\begin{aligned} c(\tilde{P}) &= \sum_i -\log(1-\rho_{i+1 \cdot 12 \dots i}^2) \\ &= -\log |\tilde{P}|, \end{aligned} \quad (3)$$

where $|\tilde{P}|$ denotes the determinant of the population correlation matrix $\tilde{P} = (\rho_{ij})$. By particularizing the result in Section 3.4

of Hsieh (1979) it is seen that the likelihood ratio test for complete independence is asymptotically optimal with the exact slope $-\log |\tilde{P}|$, the same as (3).

4. AN EMPIRICAL EVALUATION

In this section we present a Monte Carlo study of the operating characteristics of some of the asymptotic Bahadur equivalents, of the likelihood ratio test for complete independence, described in Section 3, when the samples are of moderate size. The study also includes Nagao's test given in Section 2. The finite sample behavior of the tests are investigated in terms of the power function as well as in terms of the means and s.d.'s of the P-values of the tests at various alternatives.

4.1 The Monte Carlo Experiment. The simulation study was conducted on IBM 3032 at the University of Rochester, generating the random samples from IMSL routine GGNRM. 3000 samples of size $n = 20$ and $n = 30$ were drawn from $N_p(0, \Sigma)$ with $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = 1$ and various configurations of correlations ρ_{ij} from values 0, .2, .4, .6 and .8 for $p = 3, 4, 5$. For each sample drawn, the following test statistics were obtained from the sample correlation matrix $\tilde{R} = (r_{ij})$:

- (i) Likelihood ratio based statistic $\ell = 1 - \{N-1-(2p+5)/6\} \log |\tilde{R}|$,
- (ii) Nagao's test statistic $T = (N-1) \sum_{i < j} r_{ij}^2$,
- (iii) Step-down statistics $F_i = [(N-i)/(i-1)] \cdot [R_{i \cdot 12 \dots (i-1)}^2 / (1 - R_{i \cdot 12 \dots (i-1)}^2)]$, for $i = 2, 3, \dots, p$
- (iv) Combination statistic based on Logi method

$$\psi_L = - \sum_{i=2}^p \log(P_i / (1 - P_i)),$$
- (v) Combination statistic based on Fisher's method

$$\psi_L = -2 \sum_{i=2}^p \log P_i, \text{ where } P_i \text{ are the P-values}$$

corresponding to the step down statistic.

These test statistics were compared with their critical constants determined using the following facts regarding the null distributions:

- (i) ℓ is well approximated with χ^2 distribution having d.f. $p(p-1)/2$;
- (ii) the critical constant T_α for T may be approximated by

$$T_\alpha \approx u + \frac{1}{n} \left[\frac{2a_6 u}{f(f+2)(f+4)} \{u^2 + (f+4)u + (f+2)(f+4)\} \right. \\ \left. + \frac{2a_4 \cdot u}{f(f+2)} (u+f+2) + \frac{2a_2 u}{f} \right],$$

where u is the upper 100α percentage point of the χ^2 distribution with d.f. $f = p(p-1)/2$, and a_2, a_4, a_6 are given in (1);

- (iii) F_i is distributed as a variance ratio F with $(i-1, N-i)$ d.f., for $i = 2, 3$;
- (iv) ψ_L is approximated with $a \cdot t_v$, a constant times student's t , where $a = \pi \{k(5k+2)/(3(5k+4))\}^{1/2}$, $k = p-1$, and d.f. $v = 5k+4$;
- (v) ψ_F is distributed as a χ^2 with $2(p-1)$ d.f.

The power of each test was estimated by the proportion of times the null hypothesis was rejected by the corresponding test. The s.d. of any of these estimates $\leq \{3000 \times 4\}^{-1/2}$. The P-values corresponding to the tests were obtained using the equation (1) and the results on the null distributions as mentioned above. The P-values in each case were averaged and their standard deviation was computed.

The estimated power functions and the means of the P-values of the five tests at various alternatives are given in Table 1 and Table 2 respectively. The Monte Carlo experiment with 3000 simulation was first conducted with $n = 20$ and $p = 3$ for the correlation configurations appearing in the tables. After an examination of the results it was performed with $p=4$ and 5 for two special configurations, namely (i) the extreme configuration in which only the first correlation configuration ρ_{21} is non-zero, and (ii) the symmetrical configuration where all correlation coefficients are equal. As a confirmation of the findings, the procedure was repeated with $n=30$. The s.d's of the P-values and the results for $n=30$ which are not included in this paper are available from the authors.

TABLE 1: The empirical power functions for samples of size $n=20$ with Monte Carlo of size 3000.

p	Nonzero Corr.	L.R. Test	Nagao Test	Step Dn Test	Logit Comb.	Fisher Comb.
3	H_0	.0490	.0480	.0527	.0503	.0473
	$\rho_{21} = .2$.0880	.0900	.0993	.1003	.1027
	$\phantom{\rho_{21}} = .4$.2673	.2650	.3300	.2957	.3227
	$\phantom{\rho_{21}} = .6$.6503	.6357	.7540	.6437	.7217
	$\phantom{\rho_{21}} = .8$.9823	.9780	.9930	.9717	.9877
	$\rho_{31} = .2$.0963	.0957	.0867	.0893	.0883
	$\phantom{\rho_{31}} = .4$.2710	.2663	.2420	.2197	.2347
	$\phantom{\rho_{31}} = .6$.6733	.6493	.6470	.5347	.6107
	$\phantom{\rho_{31}} = .8$.9800	.9770	.9757	.9390	.9717
	$\rho_{32} = .2$.0990	.0967	.0930	.0877	.0890
	$\phantom{\rho_{32}} = .4$.2713	.2660	.2373	.2280	.2350
	$\phantom{\rho_{32}} = .6$.6570	.6430	.6273	.5393	.5953
	$\phantom{\rho_{32}} = .8$.9717	.9650	.9707	.9210	.9587
	$\rho_{21} = \rho_{31} = .2$.1393	.1417	.1320	.1420	.1403
	$\phantom{\rho_{21} = \rho_{31}} = .4$.5983	.5400	.5270	.6200	.6170
	$\phantom{\rho_{21} = \rho_{31}} = .6$.9990	.9930	.9933	.9977	.9990
	$\rho_{31} = \rho_{32} = .2$.1387	.1357	.1187	.1173	.1220
	$\phantom{\rho_{31} = \rho_{32}} = .4$.5967	.5403	.5663	.4767	.5407
	$\phantom{\rho_{31} = \rho_{32}} = .6$.9980	.9947	.9977	.9817	.9973
	$\rho_{21} = \rho_{32} = .2$.1583	.1550	.1543	.1660	.1673
	$\phantom{\rho_{21} = \rho_{32}} = .4$.5843	.5410	.5203	.6050	.5987
	$\phantom{\rho_{21} = \rho_{32}} = .6$.9987	.9940	.9923	.9980	.9987
4	All ρ_{ij} 's = .2	.1890	.2150	.1693	.1933	.1937
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .4$.6340	.6847	.5353	.6453	.6353
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .6$.9490	.9650	.9073	.9567	.9527
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .8$	1.000	.9997	.9993	.9997	.9997
	H_0	.0473	.0457	.0457	.0480	.0463
	$\rho_{21} = .2$.0763	.0803	.0960	.0953	.1020
	$\phantom{\rho_{21}} = .4$.2013	.1937	.2817	.2350	.2707
	$\phantom{\rho_{21}} = .6$.5227	.4837	.7150	.5530	.6643
	$\phantom{\rho_{21}} = .8$.9400	.8993	.9840	.9167	.9727
	All ρ_{ij} 's = .2	.2253	.2850	.1680	.2283	.2257
5	$\phantom{\text{All } \rho_{ij} \text{'s}} = .4$.7270	.8153	.5823	.7440	.7343
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .6$.9833	.9923	.9427	.9860	.9837
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .8$	1.000	1.000	1.000	1.000	1.000
	H_0	.0470	.0463	.0447	.0527	.0473
	$\rho_{21} = .2$.0670	.0677	.0897	.0817	.0857
	$\phantom{\rho_{21}} = .4$.1520	.1493	.2517	.2043	.2313
	$\phantom{\rho_{21}} = .6$.3923	.3607	.6760	.4573	.5823
	$\phantom{\rho_{21}} = .8$.8727	.7377	.9840	.8673	.9597
	All ρ_{ij} 's = .2	.2773	.3820	.1887	.2867	.2763
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .4$.8223	.9080	.6150	.8347	.8220
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .6$.9897	.9963	.9537	.9903	.9897
	$\phantom{\text{All } \rho_{ij} \text{'s}} = .8$	1.000	1.000	1.000	1.000	1.000

TABLE 2: Estimated means of the P-values at various alternatives for samples of size 20 with Monte Carlo of size 3000.

p	Nonzero Corr.	L.R. Test	Nagao Test	Step Dn Test	Logit Comb.	Fisher Comb.
3	H_0	.4969	.4975	.4978	.4956	.4971
	$\rho_{21} = .2$.4345	.4340	.4250	.4278	.4239
	$\rho_{21} = .4$.2482	.2485	.2201	.2418	.2213
	$\rho_{21} = .6$.0775	.0805	.0555	.0863	.0610
	$\rho_{21} = .8$.0448	.0062	.0023	.0069	.0033
	$\rho_{31} = .2$.4229	.4215	.4314	.4339	.4302
	$\rho_{31} = .4$.2544	.2553	.2750	.2926	.2749
	$\rho_{31} = .6$.0793	.0823	.0900	.1234	.0968
	$\rho_{31} = .8$.0048	.0061	.0052	.0155	.0070
	$\rho_{32} = .2$.4221	.4216	.4334	.4330	.4310
	$\rho_{32} = .4$.2531	.2543	.2739	.2916	.2736
	$\rho_{32} = .6$.0834	.0832	.0919	.1248	.0975
	$\rho_{32} = .8$.0059	.0076	.0065	.0170	.0087
	$\rho_{21} = \rho_{31} = .2$.3659	.3658	.3656	.3708	.3610
	$\rho_{21} = \rho_{31} = .4$.0984	.1081	.1117	.1000	.0950
	$\rho_{21} = \rho_{31} = .6$.0007	.0035	.0028	.0008	.0007
	$\rho_{31} = \rho_{32} = .2$.3627	.3641	.3791	.3864	.3773
	$\rho_{31} = \rho_{32} = .4$.1009	.1102	.1141	.1459	.1190
	$\rho_{31} = \rho_{32} = .6$.0007	.0034	.0009	.0069	.0012
	$\rho_{21} = \rho_{32} = .2$.3606	.3613	.3607	.3637	.3562
	$\rho_{21} = \rho_{32} = .4$.1011	.1101	.1134	.1018	.0976
	$\rho_{21} = \rho_{32} = .6$.0007	.0034	.0027	.0008	.0006
	All ρ_{ij} 's = .2	.3217	.3152	.3340	.3250	.3237
	All ρ_{ij} 's = .4	.0985	.0871	.1183	.0970	.0978
	All ρ_{ij} 's = .6	.0115	.0084	.0192	.0112	.0114
	All ρ_{ij} 's = .8	.0001	.0000	.0003	.0001	.0001
4	H_0	.5030	.5017	.5037	.5064	.5047
	$\rho_{21} = .2$.4509	.4491	.4291	.4405	.4312
	$\rho_{21} = .4$.3090	.3105	.2523	.2929	.2586
	$\rho_{21} = .6$.1202	.1280	.0689	.1191	.0812
	$\rho_{21} = .8$.0124	.0204	.0031	.0177	.0057
	All ρ_{ij} 's = .2	.2876	.2705	.3145	.2910	.2904
	All ρ_{ij} 's = .4	.0660	.0484	.0981	.0643	.0653
	All ρ_{ij} 's = .6	.0040	.0018	.0125	.0034	.0039
	All ρ_{ij} 's = .8	.0000	.0000	.0002	.0000	.0000
5	H_0	.5025	.5047	.5012	.5035	.5029
	$\rho_{21} = .2$.4640	.4643	.4400	.4483	.4397
	$\rho_{21} = .4$.3353	.3346	.2626	.3056	.2724
	$\rho_{21} = .6$.1658	.1769	.0838	.1534	.1008
	$\rho_{21} = .8$.0251	.0451	.0041	.0288	.0089
	All ρ_{ij} 's = .2	.2639	.2334	.3042	.2674	.2668
	All ρ_{ij} 's = .4	.0407	.0236	.0832	.0393	.0414
	All ρ_{ij} 's = .6	.0023	.0008	.0099	.0018	.0020
	All ρ_{ij} 's = .8	.0000	.0000	.0002	.0000	.0000

4.2 *Conclusions.* Two features of the comparative behavior of the five tests clearly emerge from the two tables: (i) In the case of the extreme configuration, with $\rho_{21} \neq 0$, $\rho_{ij} = 0$ otherwise, the step down procedure is preferable. Its superiority over the other four tests increases as p increases. Nagao's test is the poorest in this case. (ii) Nagao's test dominates others if H_0 is violated in a symmetric manner, i.e., when ρ_{ij} 's are nonzero and equal. The stepdown test is the weakest in this case. It is also observed that the likelihood ratio test, and the two combinations of the P-values of the stepdown components are generally comparable and are preferable except against the two special alternatives. In view of the added resolution of the information, the authors are inclined to recommend the procedures based upon the combinations of the P-values.

REFERENCES

- Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*. Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.
- Berk, R. H., and Cohen, A. (1979). Asymptotically optimal method of combining tests. *Journal of American Statistical Association*, 74, 821-814.
- George, E. O. (1977). *Combining independent one-sided and two-sided statistical tests - some theory and applications*. Ph.D. dissertation, University of Rochester.
- Hsieh, H. K. (1979). On asymptotic optimality of likelihood ratio tests for multivariate normal distributions. *Annals of Statistics*, 7, 592-598.
- Liptak, T. (1958). On the combination of independent tests. *Magyar Tudomangos Akademia. Matematikai Kutato Intezetenek*, 3, 171-197.
- Littell, R. C. and Folks, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *Journal of American Statistical Association*, 66, 802-806.
- Mathai, A. M. and Katiyar, R. S. (1979). Exact percentage points for testing independence, *Biometrika*, 66, 353-356.
- Mudholkar, G. S. and George, E. O. (1979). The logit statistic for combining probabilities - an overview. In *Optimizing Methods in Statistics*, J. S. Rustagi, ed.
- Mudholkar, G. S. and Trivedi, M. C., and Lin, C. C. (1980). An approximation to the distribution of the likelihood ratio for complete independence. To appear in *Technometrics*.
- Nagao, H. (1973). On some test criteria for covariance matrix. *Annals of Statistics*, 1, 700-709.

- Oosterhoff, J. (1969). *Combination of One-sided Statistical Tests*, Mathematical Centre Tracts, 28, The Mathematical Centre, Amsterdam.
- Roy, S. N. and Bargmann, R. E. (1958). Tests of multiple independence and the associated confidence bounds. *Annals of Mathematical Statistics*, 29, 491-503.
- Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3, 309-326.

[Received June 1980. Revised October 1980]

ON TESTS OF INDEPENDENCE UNDER BIVARIATE EXPONENTIAL MODELS

D. R. WEIER

Department of Mathematics
and Statistics
University of South Carolina
Columbia, South Carolina 29208 USA

A. P. BASU

Department of Statistics
University of Missouri
Columbia, Missouri 65211 USA

SUMMARY. In this paper some general results are given for the asymptotic relative efficiency of Kendall's τ and Spearman's ρ under tests for independence. Discussion then turns to the bivariate exponential distribution (BVE) of Marshall and Olkin (1967) and the absolutely continuous bivariate exponential distribution (ACBVE) of Block and Basu (1974). Power calculations are carried out for several statistics, and asymptotic relative efficiencies are calculated.

KEY WORDS. tests of independence, bivariate exponential, life testing, Kendall's tau, Spearman's rho, Pitman relative efficiency, locally most powerful rank test, power study.

1. INTRODUCTION

An important topic in the area of statistical inference is the testing of independence for a pair or group of random variables. Under normality of the variables the results dealing with this topic are extensive, both in parametric and nonparametric cases. Beyond this, few other models have been available for similar analyses. Besides the general result of Blum *et al.*, (1961), some results have been attained for specially contrived

models of dependency, but these models are in general not physically appealing.

An area that is more open for investigation in testing for independence is that of distributions of the bivariate exponential type, which are more applicable as models in many physical problems, for example, in the areas of life testing and reliability analysis. Basu and Block (1975) give a good summary of distributions in this category.

In this paper we will primarily concern ourselves with the models proposed by Marshall and Olkin (1967) and Block and Basu (1974) which have appealing physical properties. The purpose of this work is the investigation of appropriate tests of independence for these models, both parametric and nonparametric, and a comparison of their powers.

We will primarily study the behavior of two well-known nonparametric tests, namely Kendall's τ and Spearman's ρ . Farlie (1960) showed that these tests are LMP and asymptotically equivalent for testing $H_0: \lambda_{12} = 0$ against $H_1: \lambda_{12} \neq 0$ when the underlying model is of the form $H(x,y) \triangleq F(x)G(y)\{1+\lambda_{12}\bar{F}(x)\bar{G}(y)\}$ where $\bar{F} = 1-F$ and $\bar{G} = 1-G$. Note that this class contains the well-known bivariate exponential distribution of Gumbel (1960). It seemed these tests would perform well for other bivariate exponential models as well.

It is well known that Kendall's τ and Spearman's ρ , after appropriate linear transformation, are asymptotically equivalent under the hypothesis of independence (see Hájek and Sídák, 1967, p. 60). If we consider contiguous alternatives of the form $\lambda_{12N} = \sigma/\sqrt{N}$, it then follows from the definition of contiguity that they are also asymptotically equivalent under contiguous alternatives. Thus $ARE(R,T) = 1$, where ARE denotes Pitman asymptotic relative efficiency.

2. TESTS FOR INDEPENDENCE

The survival function of the bivariate exponential distribution (BVE) of Marshall and Olkin (1967) is given by

$$\bar{F}(x,y) = P(X>x, Y>y) = \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x,y)],$$

where $\lambda_1, \lambda_2 > 0, \lambda_{12} \geq 0, x, y > 0$.

$$\bar{F}(x) = P(X>x) = \exp[-(\lambda_1 + \lambda_{12})x],$$

$$\bar{F}(y) = P(Y > y) = \exp[-(\lambda_2 + \lambda_{12})y].$$

Note that $\lambda_{12} = 0$ implies X and Y are independent, and since $\rho = \lambda_{12}/\lambda$ for $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, if X and Y are not independent, they must be positively correlated. Thus, a test of independence takes the form $H_0: \lambda_{12} = 0$ versus $H_1: \lambda_{12} > 0$. Now $P(X=Y)$ also equals λ_{12}/λ , so any time the sample has a diagonal element, the null hypothesis should be rejected. Thus, for a bivariate sample $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$, H_0 is rejected if $X_i = Y_i$ for any i . Since $P(X_i = Y_i) = \lambda_{12}/(\lambda_1 + \lambda_2 + \lambda_{12}) = \rho$, this event has probability $1 - (1 - \rho)^n$. For this model, $\rho \geq 0$, so for a test statistic S_n , the power function of the test is

$$1 - (1 - \rho)^n + (1 - \rho)^n P(S_n > s_0 | X \neq Y)$$

where under H_0 , $P(S_n > s_0 | X \neq Y) = \alpha$, the significance level of the test. This is because under H_0 , $\rho = 0$, and the power function becomes $P(S_n > s_0 | X \neq Y)$. Note that rejecting for diagonal elements does not contribute to the type I error. Also note that $1 - (1 - \rho)^n$ gives a lower bound on all power functions.

Bemis *et al.* (1972) derive the uniformly most powerful (UMP) test for the case when λ_1 and λ_2 are known, and Bhattacharyya and Johnson (1973) derive the UMP test for the case of identical marginals, that is, when $\lambda_1 = \lambda_2$.

The second bivariate exponential distribution to be considered is that of Block and Basu (1974) called the absolutely continuous bivariate exponential distribution (ACBVE). Its survival function $\bar{F}(x, y)$ is given by

$$\frac{\lambda}{\lambda_1 + \lambda_2} \cdot \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)] - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \cdot \exp[-\lambda \max(x, y)],$$

for $x, y > 0$, $\lambda_1, \lambda_2 > 0$, $\lambda_{12} \geq 0$, with marginals:

$$\bar{F}(x) = \frac{\lambda}{\lambda_1 + \lambda_2} \cdot \exp[-(\lambda_1 + \lambda_{12})x] - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \cdot \exp(-\lambda x),$$

$$\bar{F}(y) = \frac{\lambda}{\lambda_1 + \lambda_2} \cdot \exp[-(\lambda_2 + \lambda_{12})y] - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \cdot \exp(-\lambda y).$$

As with the BVE model, $\lambda_{12}=0$ implies X and Y are independent exponential variates with parameters λ_1 and λ_2 respectively. It is well known (see Block and Basu, 1974) that the ACBVE is the absolutely continuous component of the BVE. An immediate result of this which will be used later is that the ACBVE results when the BVE is conditioned off the diagonal. Thus when testing for independence under the bivariate exponential distribution of Marshall-Olkin and Block-Basu the power functions are related by:

$$P_{MO} = 1 - (1-\rho)^n + (1-\rho)^n P_{BB}$$

where P_{MO} and P_{BB} are the power functions of the Marshall-Olkin and Block-Basu models respectively, n denotes the sample size, and ρ the correlation of the random variables under the Marshall-Olkin model.

First consider the case of equal marginals with $\lambda_1 = \lambda_2 = \beta$ and $\lambda_{12} = \sigma$. Using the distributional results of Bhattacharyya and Johnson (1973), it can be seen that the UMP statistic U is the quotient of two independent gamma random variables. Thus, it is a simple matter to obtain the moments of U . A second statistic is based on the MLE $\hat{\delta}$ of δ as derived by Mehrotra and Michalek (1976). Moments are again easily obtained. By straightforward calculation (see Weier and Basu, 1978), one obtains the following:

Theorem 1. For the ACBVE model of Block-Basu when testing for independence with the statistics $T = \text{Kendall's } \tau$, $R = \text{Spearman's } \rho$, $U = \text{UMP statistic}$, and $M = \text{maximum likelihood statistic}$, in the identical marginal case $\text{ARE}(T,U) = \text{ARE}(R,U) = \text{ARE}(T,M) = \text{ARE}(R,M) = .5$, and $\text{ARE}(M,U) = 1$. Note that the efficiency of 1 will also hold for the BVE model due to the relationship of the two power functions.

Using the results of Shirahata (1974), the LMP rank test for the ACBVE model can be shown to be

$$S_n = \sum_{k=1}^n E_0[\max(X_{(R_k)}, Y_{(Q_k)})],$$

which, after some simplification, reduces to the statistic

$$L = \sum_{k=1}^n \sum_{i=0}^{R_k-1} \sum_{j=0}^{Q_k-1} \frac{\binom{n}{i} \binom{n}{j}}{\binom{2n}{i+k} \binom{2n-1-j}{2n-1-j}} .$$

For details see Weier and Basu (1978).

In several instances the statistic giving the LMPRT is of the form where the argument, say $A_n(R_k, Q_k)$, factors into $A_n(R_k) B_n(Q_k)$. Examples are the normal scores statistic and Spearman's ρ . Although the ACBVE model does not give rise to such a statistic, the presence of exponentiality would make it seem of interest to check the performance of such a statistic where A_n and B_n are the expected values of order statistics from an exponential distribution. We then have

$$S = \sum_{k=1}^n A_n(R_k) B_n(Q_k) = \sum_{k=1}^n \left[\sum_{i=n-R_k-1}^n 1/i \cdot \sum_{j=n-Q_k-1}^n 1/j \right] .$$

Here S will be called the exponential scores statistic. Suitability of an exponential score statistic for a life testing problem was studied by Basu (1968). All the statistics, except L and S , are known to be asymptotically normally distributed. We next compare their powers.

3. POWER COMPARISONS

Large sample power calculations are given in Table 1. Recall that the power functions for the BVE and ACBVE models are related by $P_{MO} = 1 - (1-\rho)^n + (1-\rho)^n P_{BB}$ where $P_{MO}(P_{BB})$ is the power for the BVE (ACBVE) distribution, and $\rho = \lambda_{12}/\lambda = \delta/(2\beta+\delta)$, the correlation coefficient for the BVE distribution. The tabled values are the probabilities of rejection when α is the significance level, and ρ indicates the value of the correlation coefficient under the alternatives. Recall that this is a one-sided test. χ^2 indicates the power of the statistic proposed by Bemis *et al.* (1972) in the case where λ_1 and λ_2 are known. The values are exact chi-square probabilities, and they provide upper bounds on powers for both models. U provides exact F-probabilities for the UMP test for Bhattacharyya and Johnson (1973). T indicates the normal approximation to the distribution of Kendall's τ . r indicates powers obtained using the normal approximation to the distribution of the usual

TABLE 1: Large sample power of tests.

ρ	$\alpha=0.05$	ACBVE				BVE			
		χ^2	U	T	r	χ^2	U	T	r
0.01	10	0.0538	0.0522	0.0514	0.0516	0.1443	0.1429	0.1421	0.1423
	20	0.0554	0.0533	0.0521	0.0523	0.2274	0.2256	0.2247	0.2249
	40	0.0575	0.0547	0.0532	0.0533	0.3695	0.3676	0.3666	0.3667
	80	0.0591	0.0568	0.0546	0.0548	0.5789	0.5779	0.5769	0.5770
	160	0.0640	0.0598	0.0568	0.0569	1.0000	1.0000	1.0000	1.0000
0.05	10	0.0712	0.0617	0.0572	0.0587	0.4439	0.4382	0.4355	0.4364
	20	0.0804	0.0676	0.0613	0.0626	0.6703	0.6657	0.6635	0.6640
	40	0.0946	0.0765	0.0673	0.0685	0.8837	0.8813	0.8801	0.8803
	80	0.1172	0.0905	0.0764	0.0776	0.9854	0.9850	0.9847	0.9848
	160	0.1571	0.1132	0.0906	0.0919	1.0000	1.0000	1.0000	1.0000
0.10	10	0.0968	0.0749	0.0649	0.0685	0.6851	0.6774	0.6739	0.6752
	20	0.1199	0.0886	0.0741	0.0775	0.8930	0.8892	0.8874	0.8878
	40	0.1583	0.1107	0.0882	0.0918	0.9876	0.9869	0.9865	0.9866
	80	0.2263	0.1481	0.1110	0.1153	0.9998	0.9998	0.9998	0.9998
	160	0.3439	0.2135	0.1493	0.1553	1.0000	1.0000	1.0000	1.0000
0.30	10	0.2352	0.1403	0.1004	0.1190	0.9784	0.9757	0.9746	0.9751
	20	0.3484	0.2037	0.1390	0.1618	0.9995	0.9994	0.9993	0.9993
	40	0.5308	0.3157	0.2064	0.2375	1.0000	1.0000	1.0000	1.0000
	80	0.7705	0.5038	0.3258	0.3710	1.0000	1.0000	1.0000	1.0000
	160	0.9511	0.7576	0.5244	0.5855	1.0000	1.0000	1.0000	1.0000
0.60	10	0.4809	0.2661	0.1650	0.2243	0.9999	0.9999	0.9999	0.9999
	20	0.7011	0.4297	0.2684	0.3482	1.0000	1.0000	1.0000	1.0000
	40	0.9094	0.6721	0.4481	0.5515	1.0000	1.0000	1.0000	1.0000
	80	0.9924	0.9065	0.7065	0.8064	1.0000	1.0000	1.0000	1.0000
	160	1.0000	0.9946	0.9310	0.9715	1.0000	1.0000	1.0000	1.0000

TABLE 1 (cont.).

$\alpha=0.10$ ρ		ACBE					BVE				
		χ^2	U	T	r		χ^2	U	T	r	
0.01	10	0.1063	0.1038	0.1023	0.1027		0.1917	0.1895	0.1881	0.1885	
	20	0.1089	0.1055	0.1036	0.1039		0.2712	0.2684	0.2668	0.2671	
	40	0.1123	0.1080	0.1053	0.1056		0.4062	0.4033	0.4015	0.4017	
	80	0.1167	0.1115	0.1078	0.1080		0.6047	0.6024	0.6007	0.6008	
	160	0.1244	0.1165	0.1113	0.1115		1.0000	1.0000	1.0000	1.0000	
0.05	10	0.1335	0.1198	0.1120	0.1145		0.4812	0.4730	0.4683	0.4698	
	20	0.1481	0.1293	0.1188	0.1209		0.6946	0.6879	0.6841	0.6849	
	40	0.1697	0.1434	0.1286	0.1305		0.8933	0.8899	0.8880	0.8883	
	80	0.2047	0.1650	0.1430	0.1449		0.9869	0.9862	0.9858	0.9859	
	160	0.2594	0.1988	0.1649	0.1670		1.0000	1.0000	1.0000	1.0000	
0.10	10	0.1714	0.1413	0.1246	0.1305		0.7111	0.7006	0.6948	0.6968	
	20	0.2051	0.1624	0.1394	0.1448		0.9034	0.8982	0.8954	0.8960	
	40	0.2580	0.1954	0.1614	0.1669		0.9890	0.9881	0.9876	0.9877	
	80	0.3462	0.2481	0.1953	0.2016		0.9999	0.9998	0.9998	0.9998	
	160	0.4804	0.3335	0.2493	0.2575		1.0000	1.0000	1.0000	1.0000	
0.30	10	0.3483	0.2390	0.1798	0.2070		0.9816	0.9785	0.9768	0.9776	
	20	0.4772	0.3223	0.2353	0.2663		0.9996	0.9995	0.9994	0.9994	
	40	0.6591	0.4546	0.3243	0.3626		1.0000	1.0000	1.0000	1.0000	
	80	0.8571	0.6458	0.4647	0.5135		1.0000	1.0000	1.0000	1.0000	
	160	0.9762	0.8560	0.6643	0.7187		1.0000	1.0000	1.0000	1.0000	
0.60	10	0.6033	0.4004	0.2705	0.3465		1.0000	0.9999	0.9999	0.9999	
	20	0.7990	0.5757	0.3995	0.4891		1.0000	1.0000	1.0000	1.0000	
	40	0.9503	0.7913	0.5920	0.6888		1.0000	1.0000	1.0000	1.0000	
	80	0.9968	0.9538	0.8176	0.8902		1.0000	1.0000	1.0000	1.0000	
	160	1.0000	0.9982	0.9676	0.9883		1.0000	1.0000	1.0000	1.0000	

product-moment correlation estimator. The convergence to normality of this statistic is slow under normal alternatives, and the moments used are those obtained under normal assumptions, so the reliability of the values obtained here for the exponential models is questionable.

As should be the case, other than the χ^2 test, the UMP statistic has the best performance throughout. r performs somewhat better than T , but again the moments and convergence to normality are appropriate properties under normality but are questionable in this situation. Subsequent small sample studies did tend to bear out this result however. It can be noted throughout the table that in the ACBVE case, for T to achieve equivalent power with respect to U , the sample size must be doubled. For example, consider $\alpha=.10$ and $\rho=.05$; for $n=80$ the power for U is .1650 while for $n=160$ the power for T is .1649. This provides evidence supporting the earlier result that $ARE(T,U) = .5$.

For small sample considerations, the null distributions of the LMPRT statistic L and the exponential scores statistic S were generated for $n = 4, 5$, and 6 . For fixed increasing X ranks, all permutations of the Y ranks were generated, and values of the statistic were computed in each case. For $n \geq 7$ the magnitude of $n!$ begins to prohibit further generation of the null distributions. Selected quantiles of the computed distributions are given in Table 2. Note that positive correlations is indicated by large values of both statistics so independence would be rejected in favor of positive correlation when values of the statistic exceed the appropriate upper quantiles.

The results of a small sample Monte Carlo study are given in Table 3. The tabled values are the number of rejections out of 10,000 trials for $n=6$ with α (the significance level) = .10, and $\beta=1$. U , L , and S are as used previously. T and R denote Kendall's τ and Spearman's ρ respectively. r again represents the usual product moment estimate of correlation; r_t yields the number of rejections obtained when using the .9 quantile of the usual t statistic for r which is appropriate for data which is normally distributed; and r_s is the number of rejections obtained using the .9 quantile of a previous simulation of ACBVE data based on 10,000 trials. Exact critical regions of size .10 were obtained for L , S , T , and R by randomization on the .9 quantiles.

The section of the table labelled ACBVE consists of power of the statistics for data generated under the ACBVE model. The generation scheme used was that described in Friday and Patil (1977), page 543.

TABLE 2: Null distributions of score statistics. For given α , S_α and L_α are the values such that $P(S \leq S_\alpha) = \alpha$ and $P(L \leq L_\alpha) = \alpha$.

n =	4		5		6	
α	L_α	S_α	L_α	S_α	L_α	S_α
.01	1.4619	2.3055	1.7917	2.6819	2.1440	3.3017
.05	1.5191	2.5555	1.8709	3.0986	2.2300	3.7267
.10	1.5762	2.6389	1.9306	3.2653	2.3655	4.0517
.20	1.7333	2.9722	2.1091	3.6819	2.5218	4.5086
.30	1.7905	3.0555	2.2480	4.1680	2.6858	4.9336
.40	1.8762	3.3889	2.3472	4.4597	2.8323	5.3795
.50	1.9333	3.6666	2.4464	4.7514	2.9549	5.7514
.60	2.1143	3.8889	2.5853	5.1889	3.1089	6.1681
.70	2.2905	4.8055	2.8837	5.5430	3.2655	6.6889
.80	2.3619	5.2222	2.9424	6.5569	3.4764	7.7375
.90	2.5429	5.3889	3.1210	7.0222	3.7061	8.3975
.95	2.5762	5.6666	3.2599	7.4041	3.9109	8.8556
.99	2.6714	5.8055	3.4682	7.6055	4.1266	9.3000

TABLE 3: Small sample power.

ACBVE		U	L	S	T	R	r_s	r_t
λ_{12}								
	ρ							
0.0	0.0	1025	1026	1002	1043	1050	1049	1248
0.506	0.1	1606	1380	1362	1362	1315	1363	1602
1.414	0.2	2413	1895	1860	1788	1774	1966	2246
21.489	0.4	4221	2879	2809	2752	2737	3128	3526
Weibull								
	ρ							
0.5	0.0	6126	993	960	1000	988	798	978
	0.2	8056	1889	1826	1763	1748	1651	2019
1.5	0.0	350	1021	1014	974	988	1158	1322
	0.2	1092	1911	1840	1762	1777	2156	2518
Normal								
	ρ							
0.0		4	981	974	1006	1009	795	1002
0.3		1935	2247	2182	2247	2286	2106	2518
Lognormal								
	ρ							
0.0		1098	957	978	1002	996	1066	1248
0.3		2143	2225	2085	2289	2301	2299	2613

The values of ρ in this section refer to the correlation coefficient of the ACBVE distribution, not that of the BVE as in the large sample study. For the ACBVE, $\rho = (4\delta\beta + 3\delta^2) / (16\beta^2 + 20\delta\beta + 7\delta^2)$ [see Block and Basu, 1974, letting $\lambda_1 = \lambda_2 = \beta$, $\lambda_{12} = \delta$]. Note that the maximum value occurs for $\delta \rightarrow \infty$, and it is $3/7$ or, approximately, .4286. The simulation was not done for the BVE model due to the direct relationship of the two models' power functions.

To study robustness of the tests, a number of bivariate distributions are considered. The Weibull section is simply bivariate data from Weibull distributions generated in the manner of the ACBVE distribution with correlation ρ and then raised to power m . The normal section is based on bivariate normal data with means and variance of one and correlation ρ . The lognormal uses pairs (e^X, e^Y) where (X, Y) is generated as in the normal case.

Except for r_t the number of rejections under H_0 which should be 1000 is acceptable in all cases. For the data of the ACBVE type, the UMP test certainly performs the best. As is expected, R and T perform comparably throughout. The inappropriateness of using the t distribution to obtain the critical region for r can be seen since the type I error is too large, thus exaggerating the power throughout. With the more appropriate modified critical region, r is second in performance to the UMP test. However, in general its distribution and thus this critical region would not be known. The two score statistics do better than the usual two rank statistics with the LMPRT performing better than the exponential scores. However, it is disappointing that they do not compare more favorably with the UMP test. If this performance carries through to large sample power, one would suspect that perhaps the score statistics are asymptotically as efficient as the other rank tests and not the UMP test. Since $ARE(T, U) = ARE(R, U) = .5$, we do know that $.5 \leq ARE(L, U) \leq 1$, and the results suggest it may be closer to .5 than 1, not at all the situation with the bivariate normal case concerning the normal scores statistic.

Since rank tests are invariant under order-preserving transformations their performance remains the same in the Weibull case as in the ACBVE. Again the scores statistics do better than the other two with the LMPRT the best. The parametric statistics are quite distorted with respect to type I error, and thus their values are of little consequence since they are highly non-robust.

In the normal case r_t is now uniformly most powerful, and it performs as such. r_2 is no longer of importance since $P(\text{Type I}) < .1$. Now, as would be expected, R and T perform more favorably than the score statistics based on exponential assumptions. The same carries through to the lognormal case as again the orderings of the data have been preserved. Here r computed for the pairs $(\ln X, \ln Y)$ would be uniformly most powerful with numbers of rejections approximately those of r_t in the normal case.

It would seem from the information in this section that although rank statistics perform well under assumptions of normality, the same does not appear to be true under this type of exponential model. Rank statistics are quite lacking in efficiency when compared to their parametric counterparts. However, note that this means for the rank statistics to achieve the same power as the parametric, a much larger sample size is required; when one considers a given sample size (see Tables 1 and 3), the resulting powers are not as radically different as the efficiency results might indicate. Also, one has to be extremely careful about choosing the right parametric model since the parametric statistics are highly non-robust. Though of lower efficiency, all the rank tests are seen to be quite robust as to model variations.

Additional results for multivariate exponential distributions have been obtained (see Weier and Basu, 1981) and will appear elsewhere.

This research has been supported in part under contract N00014-78-C-0655 for the Office of Naval Research.

REFERENCES

- Basu, A. P. (1968). On a generalized savage statistic with applications to life testing. *Annals of Mathematical Statistics*, 39, 1591-1604.
- Basu, A. P. and Block, H. W. (1975). On characterizing univariate and multivariate exponential distributions with applications. In *Statistical Distributions in Scientific Work*, Vol. 3, G. P. Patil, S. Kotz, and J. K. Ord, eds. Reidel, Dordrecht-Holland. Pages 399-421.
- Bemis, B. M., Bain, L. J., and Higgins, J. J. (1972). Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. *Journal of the American Statistical Association*, 67, 927-929.
- Bhattacharyya, G. L. and Johnson, R. A. (1973). On a test of independence in a bivariate exponential distribution. *Journal of the American Statistical Association*, 68, 704-706.

- Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69, 1031-1037.
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32, 485-498.
- Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47, 307-323.
- Friday, D. S. and Patil, G. P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variables. In *The Theory and Applications of Reliability*, G. P. Tsokos and I. N. Shimi, eds. Academic Press, New York. Pages 527-549.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698-707.
- Hájek, J. and Sidák, Z. (1967). *The Theory of Rank Tests*. Academic Press, New York.
- Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.
- Mehrotra, K. G. and Michalek, J. E. (1976). Estimation of parameters and tests of independence in a continuous bivariate exponential distribution. (Unpublished.)
- Shirahata, S. (1974). Locally most powerful rank tests for independence. *Bulletin of Mathematical Statistics*, 16, 11-21.
- Weier, D. R. and Basu, A. P. (1978). On tests of independence under bivariate exponential models. University of Missouri-Columbia, Department of Statistics, Technical Report No. 76.
- Weier, D. R. and Basu, A. P. (1981). Testing for independence in multivariate exponential distributions. To appear in *Australian Journal of Statistics*.

[Received May 1980. Revised September 1980]

ON TESTS FOR DETECTING CHANGE IN THE MULTIVARIATE MEAN

M. S. SRIVASTAVA¹

Department of Statistics
University of Toronto
Toronto, Canada M5S 1A1

SUMMARY. We consider tests based on one observation on each of $N \geq 2$ random vectors $\underline{x}_1, \dots, \underline{x}_N$ to decide if the mean vectors $\underline{\mu}_i$ of the \underline{x}_i 's are all equal against the alternative that a change has occurred at some unknown point r , (i.e., $\underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_r \neq \underline{\mu}_{r+1} = \dots = \underline{\mu}_N$). The \underline{x}_i 's are assumed to be normally distributed with common unknown covariance. An estimate of the change point r is also given.

KEY WORDS. detecting a change point, multivariate mean, likelihood ratio test and estimate, distribution of test statistics.

1. INTRODUCTION

A problem of considerable practical interest is the following: Given one observation from each of N random vectors $\underline{x}_1, \dots, \underline{x}_N$, how can we decide whether the means of the \underline{x}_i 's can be considered to be the same or whether one needs to consider two models of the form

$$\underline{x}_i = \underline{\mu} + \underline{\varepsilon}_i \quad (1 \leq i \leq r), \quad \underline{x}_i = \underline{\mu}^* + \underline{\varepsilon}_i \quad (r+1 \leq i \leq N)$$

where the $\underline{\varepsilon}_i$'s are independent error p -vectors and r is unknown.

¹Research supported by National Research Council of Canada Grant No. A4060.

Apart from obvious potential applications to the detection of shifts in production processes, tests of this kind may be applied to the detection of impacts of social programs (drugs, advertising campaign) since the time at which the effects of the program are felt is usually unknown. When $p = 1$ and the initial level (μ_1) is known, this problem was first considered by Page (1955) for one-sided alternatives. Chernoff and Zacks (1964) gave a Bayes test which was later generalized by Gardner (1969) to two-sided alternatives; the exact distribution along with percentage points were given by Sen and Srivastava (1975a). The unknown variance case was considered by Sen and Srivastava (1975b). Sen and Srivastava (1975a,b,c) also proposed likelihood ratio tests and compared its power with Bayes tests. If there has been only one change and the change occurred near the end, the likelihood ratio tests perform better than the Bayes test. Bhattacharya and Johnson (1968), Sen and Srivastava (1975a) and Sen (1971) gave some nonparametric tests. However, when $p > 1$, Sen and Srivastava proposed a Bayes test when the covariance matrix Σ is of the form $\Sigma = \sigma^2 I$, σ^2 unknown.

In this paper, we propose an estimate of 'r', the change point and two tests for detecting the change point in the multivariate mean when the covariance matrix is completely unknown. While the percentage points are available for one of the tests, we provide tables for the second test up to $p = 5$ by Monte Carlo techniques.

The problem of detecting the change and the problem of estimating the change point can be stated formally as follows:

Problem. Let $\underline{x}_1, \dots, \underline{x}_N$ be independently normally distributed as $N_p(\underline{\mu}_i, \Sigma)$, $i = 1, 2, \dots, N$, $\Sigma > 0$. The problem is to test the hypothesis

$$H : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_N \quad (1)$$

vs.

$$A : \underline{\mu}_1 = \dots = \underline{\mu}_r \neq \underline{\mu}_{r+1} = \dots = \underline{\mu}_N \quad (2)$$

where the point of change 'r' is not known. If the change has occurred, an estimate of the change point 'r' is desired.

2. THE LIKELIHOOD METHOD

Let $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$ be independently distributed as $N_p(\underline{\mu}_i, \Sigma)$, $i = 1, 2, \dots, N$, $\Sigma > 0$. Let

$$X = (\underline{x}_1, \dots, \underline{x}_N), \quad \bar{\underline{x}}_r = r^{-1} \sum_{i=1}^r \underline{x}_i, \quad \bar{\underline{x}}_{N-r} = (N-r)^{-1} \sum_{i=r+1}^N \underline{x}_i,$$

$$W_r = \sum_{i=1}^r (\underline{x}_i - \bar{\underline{x}}_r)(\underline{x}_i - \bar{\underline{x}}_r)' + \sum_{i=r+1}^N (\underline{x}_i - \bar{\underline{x}}_{N-r})(\underline{x}_i - \bar{\underline{x}}_{N-r})',$$

$$\underline{y}_r = N^{-\frac{1}{2}} r^{\frac{1}{2}} (N-r)^{\frac{1}{2}} (\bar{\underline{x}}_r - \bar{\underline{x}}_{N-r}), \quad T_r = \underline{y}_r' W_r^{-1} \underline{y}_r,$$

$$\text{and} \quad T_{\hat{r}} = \max_{1 \leq r \leq N-1} T_r, \quad (3)$$

where \hat{r} is the point where the maximum occurs. Then \hat{r} is the maximum likelihood estimate of r and $T_{\hat{r}}$ is the test statistic for testing the hypothesis H in (1) vs A in (2).

2.1 Distribution of the Test Statistic $T_{\hat{r}}$. The distribution of the statistic $T_{\hat{r}}$ or equivalently $S_{\hat{r}}$ is difficult to obtain. The percentage points given in Table 1 are obtained by Monte Carlo techniques. For the benefit of the readers we have reproduced in the last portion of the table the percentage points for one-sided test given in Sen and Srivastava (1975c).

In order to test the accuracy of these tables, we give below an upper and lower bound for the percentage points by Bonferroni inequality:

$$P_{2r+1} \leq P\{E_1 \cap E_2 \cap \dots \cap E_k\} \leq P_{2r+2}$$

$$\text{where } Pr = 1 - \sum_{i=1}^k P(E_i^c) + \sum_{i_1 < i_2} P(E_{i_1}^c \cap E_{i_2}^c) - \dots$$

$$+ (-1)^r \sum_{i_1 < \dots < i_r} P(E_{i_1}^c \cap E_{i_2}^c \cap \dots \cap E_{i_r}^c),$$

E_1, E_2, \dots, E_k are events and E_i^c is the complementary events of E_i .

Since the statistic T_r is invariant under nonsingular linear transformations, we shall assume without loss of generality that $\Sigma = I$. Let

TABLE 1: Percentage points of $P\{S_r^* > c\}$. Percentage points for the one-sided test with $p = 1$ taken from Sen and Srivastava (1975c).

n	90%	95%	99%	90%	95%	99%
p = 1			p = 2			
9	0.59850	0.67943	0.79930	0.75688	0.81341	0.89097
10	0.54074	0.62353	0.74597	0.71104	0.77004	0.85087
12	0.47743	0.55000	0.67171	0.62747	0.68406	0.78988
15	0.40337	0.46775	0.59305	0.53145	0.58507	0.68743
20	0.32205	0.37556	0.48757	0.42853	0.48202	0.57685
25	0.26561	0.31068	0.40067	0.35524	0.39723	0.49273
30	0.22703	0.26905	0.35556	0.30784	0.34571	0.43358
35	0.20006	0.23469	0.30356	0.26572	0.30244	0.38223
40	0.17682	0.20983	0.27986	0.23835	0.26818	0.35599
45	0.142983	0.17122	0.22893	0.21333	0.24203	0.30162
50	0.14605	0.17122	0.22893	0.19443	0.22079	0.27842
55	0.13435	0.14865	0.21530	0.18075	0.20658	0.25332
60	0.12273	0.14556	0.19371	0.16712	0.19042	0.24013
p = 3			p = 4			
9	0.86321	0.90100	0.95274	0.93122	0.95225	0.98030
10	0.81646	0.86036	0.91772	0.89216	0.92050	0.95949
12	0.73506	0.78009	0.86278	0.81345	0.85251	0.90931
15	0.62586	0.67883	0.76367	0.70757	0.75129	0.82921
20	0.50618	0.55494	0.64622	0.57555	0.62176	0.69995
25	0.42409	0.46583	0.54144	0.48724	0.53224	0.49789
30	0.36786	0.40867	0.48409	0.41876	0.45786	0.53528
35	0.32011	0.35600	0.42975	0.37243	0.40890	0.48475
40	0.28631	0.31873	0.38936	0.32978	0.36244	0.43006
45	0.25956	0.28885	0.35059	0.29773	0.32898	0.38657
50	0.23737	0.26597	0.32541	0.27333	0.30196	0.35932
55	0.21703	0.24360	0.30658	0.25160	0.27796	0.33059
60	0.20169	0.22531	0.27546	0.23334	0.25907	0.31084
p = 5			p = 1, one-sided test			
9	0.97470	0.98470	0.99501	2.68766	3.26172	4.6487
10	0.94582	0.96234	0.98435	2.58329	3.08961	4.32861
12	0.87772	0.90674	0.94571	2.57171	3.01776	4.08071
15	0.77460	0.81265	0.87580	2.51132	2.94701	3.90726
20	0.63872	0.68443	0.75851	2.51568	2.90492	3.76205
25	0.54015	0.58265	0.65908	2.48975	2.86356	3.59018
30	0.46803	0.50539	0.57951	2.49352	2.85028	3.57691
35	0.41320	0.44663	0.51773	2.53353	2.85028	3.57691
40	0.36919	0.40287	0.46901	2.52323	2.85604	3.48837
45	0.33413	0.36387	0.42353	2.50628	2.83838	3.48449
50	0.30807	0.33579	0.39341	2.54559	2.86174	3.50783
55	0.28220	0.30740	0.36533	2.53852	2.85646	3.54822
60	0.26006	0.28626	0.33945	2.55428	2.85276	3.46943

$$S_r = y_r' V^{-1} y_r, \text{ where } V = \sum_{i=1}^N (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})',$$

$$\text{and } S_r^\wedge = \max_{1 \leq r \leq N-1} S_r, \quad \bar{\underline{x}} = N^{-1} \sum_{i=1}^N \underline{x}_i. \quad (4)$$

Then $T_r = S_r(1 - S_r)^{-1}$ and the maximum likelihood test is equivalent to the one based on S_r^\wedge . And the maximum likelihood estimate of r can be obtained either from (3) or (4). We shall now obtain a lower and an upper bound for $P\{S_r^\wedge > c_\alpha\}$. Let

$$\underline{a}'_r = \underbrace{N^{-\frac{1}{2}}((N-r)^{\frac{1}{2}}r^{-\frac{1}{2}}, \dots, (N-r)^{\frac{1}{2}}r^{-\frac{1}{2}})}_{r \text{ terms}}, \underbrace{-(N-r)^{-\frac{1}{2}}r^{\frac{1}{2}}, \dots, -(N-r)^{-\frac{1}{2}}r^{\frac{1}{2}}}_{(N-r) \text{ terms}}$$

Then $S_r = \underline{a}'_r X' V^{-1} X \underline{a}_r$. Note that $\underline{a}'_r \underline{e} = 0$, where $\underline{e}' = (1, \dots, 1) : 1 \times N$, and $\underline{a}'_r \underline{a}_r = 1$. Hence, from the results of Section 3.1,

$$\frac{N-p-1}{p} \frac{S_r}{1-S_r} \sim F_{p, N-p-1}$$

where ' \sim ' denotes 'distributed like' and $F_{m,n}$ denotes an F-distribution with (m,n) degrees of freedom (d.f.).

We shall now obtain the joint distribution of S_ℓ and S_t , $\ell \neq t$. Let

$$A_{\ell t} = \begin{bmatrix} \underline{a}'_\ell \\ \underline{a}'_t \end{bmatrix} : 2 \times N \text{ and } M = I - N^{-1} \underline{e} \underline{e}',$$

where $\underline{e}' = (1, 1, \dots, 1) : 1 \times N$. For the sake of notational convenience, we shall drop the subscript ' ℓt ' from $A_{\ell t}$.

Let C be a $(N-3) \times N$ matrix of rank $(N-3)$ such that $C \underline{e} = \underline{0}$ and $CA' = 0$; note that $A \underline{e} = 0$. Then

$$V = XMX' = XA'(AA')^{-1}AX' + XC'(CC')^{-1}CX'.$$

Hence, XA' and XC' are independently distributed. In the notation of Srivastava and Khatri (1979, pp. 170-171),

$XA' \sim N_{p,2}(0, I, AA')$ and $XC'(CC')^{-1}CX' \sim W_p(I, N-3)$ for

$p \leq N-3$. We shall assume that $p \leq N-3$, $p \geq 2$. Let $U = XA'$ and $W = XC'(CC')^{-1}CX$. Then the joint pdf of U and W is given by

$$c |AA'|^{-\frac{1}{2}p} |W|^{\frac{1}{2}(N-p-4)} \text{etr} - \frac{1}{2}[U(AA')^{-1}U' + W]$$

where $c = [2^{\frac{1}{2}p(N-3)} \Gamma_p(\frac{1}{2}(N-3))]^{-1}$. Let

$$V = W + U(AA')^{-1}U' \quad \text{and} \quad V^{-\frac{1}{2}}U = R.$$

Then the joint pdf of R and V is given by

$$c |AA'|^{-\frac{1}{2}p} |V|^{\frac{1}{2}(N-p-2)} |I - (AA')^{-1}R'R|^{\frac{1}{2}(N-p-4)} \text{etr} - \frac{1}{2}V.$$

Hence, the pdf of R is given by

$$c_1 |AA'|^{-\frac{1}{2}p} |I - (AA')^{-1}R'R|^{\frac{1}{2}(N-p-4)},$$

where $c_1 = 2^p \Gamma(\frac{N-1}{2}) \Gamma(\frac{N-2}{2}) / \Gamma(\frac{N-p-1}{2}) \Gamma(\frac{N-p-2}{2})$. Let $G = R'R$. Then,

the pdf of G (see Srivastava and Khatri 1979, Lemma 3.2.3, p. 76) is given by

$$c_2 |B|^{-\frac{1}{2}p} |G|^{\frac{1}{2}(p-3)} |I - B^{-1}G|^{\frac{1}{2}(N-p-4)},$$

where $B = AA'$, and

$$c_2 = 2^p \pi^p \Gamma(\frac{N-1}{2}) \Gamma(\frac{N-2}{2}) / \Gamma(\frac{N-p-1}{2}) \Gamma(\frac{N-p-2}{2}) \Gamma(\frac{p}{2}) \Gamma(\frac{p-1}{2}).$$

Letting $g_0 = \underline{a}'_l X' V^{-1} X \underline{a}_t$ we find that

$$G = \begin{bmatrix} S_\ell & g_0 \\ g_0 & S_t \end{bmatrix}$$

Hence, to get the joint pdf of S_ℓ and S_t , we need to integrate out g_0 . Expanding the determinants in the above expression for the pdf of G , we get the joint pdf of S_ℓ , S_t and g_0 given by

$$c_2 |B|^{-\frac{1}{2}p} (S_\ell S_t - g_0^2)^{\frac{1}{2}(p-3)} [1 - b(S_\ell + S_t) + (b^2 - d^2)S_\ell S_t - 2dg_0 + (d^2 - b^2)g_0^2]^{\frac{1}{2}(N-p-4)}$$

where $B^{-1} \equiv \begin{bmatrix} b & d \\ d & b \end{bmatrix}$, $S_\ell S_t - g_0^2 > 0$ and $|I - B^{-1}G| > 0$.

Expanding the last expression given in the bracket [], term by term integration can be carried out and the joint pdf of S_ℓ and S_t can be obtained. For computational simplicity, we may choose N and p such that $(N-p-4)/2$ is an integer. Thus

$$1 - \sum_{i=1}^{N-1} P(S_i > c_\alpha) \leq 1 - P\{\hat{S}_r > c_\alpha\} \leq 1 - \sum_{i=1}^{N-1} P(S_i > c_\alpha) \\ + \sum_{i_1 < i_2}^{N-1} P(S_{i_1} > c_\alpha, S_{i_2} > c_\alpha).$$

That is,

$$\sum_{i=1}^{N-1} P(S_i > c_\alpha) - \sum_{i_1 < i_2}^{N-1} P(S_{i_1} > c_\alpha, S_{i_2} > c_\alpha) \\ \leq P(\hat{S}_r > c_\alpha) \leq \sum_{i=1}^{N-1} P(S_i > c_\alpha).$$

These upper and lower bounds of the percentage points may be used to check the accuracy of the simulation result given in Table 1. If these bounds or some linear combination of it (say average) is close to the simulation result, then they may be used to obtain tables for many more values of p . However, this has not been done here. This along with power comparisons of the two procedures are planned for a future communication. It may be noted that in all of the studies carried out by Sen and Srivastava (1975a,b,c, 1973), the likelihood ratio test seems to have a better power than other competing tests when the change point occurs near the end (a more practical situation). Also, the likelihood method gives an estimate of the change point.

3. AN ALTERNATIVE TEST

Let $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$ be independently distributed random p -vectors. Define

$$\underline{u} = \sum_{i=1}^{N-1} i(\underline{x}_{i+1} - \bar{\underline{x}})/(\underline{a}'\underline{a})^{1/2} \equiv (u^{(1)}, \dots, u^{(p)})',$$

where $\underline{a}' = \frac{1}{2}[-(N-1), -(N-3), -(N-5), \dots, (N-5), (N-3), (N-1)]$

and $\bar{\underline{x}} = N^{-1} \sum_{i=1}^N \underline{x}_i$. Let $V = \sum_{i=1}^N (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$. Then we

propose a test statistic

$$Q = \underline{u}' V^{-1} \underline{u}$$

for testing the hypothesis H against the alternative A . It may be noted that $u^{(j)}$ is the univariate test statistic used by Chernoff and Zacks (1964) for testing one-sided change when variance is known. It should be mentioned that a nonparametric generalization of the Bhattacharya and Johnson (1968) test can also be given on this line.

3.1 Distribution of the Statistic Q. Let X denote the observation matrix, $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N) : p \times N$. Then, it can easily be shown that

$$\underline{u}' = \underline{b}' X' \quad \text{and} \quad V = XX' - N \underline{\bar{x}} \underline{\bar{x}}', \quad \underline{b} = \underline{a} / (\underline{a}' \underline{a})^{\frac{1}{2}}$$

where $\underline{\bar{x}} = N^{-1} \sum_{i=1}^N \underline{x}_i$, and the vector \underline{a} has been defined above.

Thus

$$Q = \underline{b}' X' V^{-1} X \underline{b} \quad \text{and} \quad 1 + Q = |V + X \underline{b} \underline{b}' X'| / |V|$$

where $|B|$ denotes the determinant of the matrix B . Let Γ be an orthogonal matrix of the order $N \times N$ whose first row is given by $\underline{e}' / N^{\frac{1}{2}}$, and the second row is given by $\underline{a}' / (\underline{a}' \underline{a})^{\frac{1}{2}}$, where $\underline{e}' = (1, 1, \dots, 1) 1 \times N$, a row N -vector of ones. That is

$$\Gamma = \begin{bmatrix} \underline{e}' / N^{\frac{1}{2}} \\ \underline{a}' / (\underline{a}' \underline{a})^{\frac{1}{2}} \\ C \end{bmatrix}$$

where C is an $(N-2) \times N$ matrix such that Γ is an orthogonal matrix.

Let $Y = (y_1, \dots, y_N) = X \Gamma'$. Then

$$V = \sum_{i=1}^N y_i y_i' \quad \text{and} \quad X \underline{b} \underline{b}' X' = Y \Gamma \underline{b} \underline{b}' \Gamma' Y = y_2 y_2',$$

and under the hypothesis H , y_2, \dots, y_N are independently distributed as $N_p(0, \Sigma)$. And

$$1 + Q = \frac{|2y_2 y_2' + \sum_3^N y_i y_i'|}{|y_2 y_2' + \sum_3^N y_i y_i'|} = \frac{1 + 2y_2' W^{-1} y_2}{1 + 2y_2' W^{-1} y_2},$$

where $W = \sum_3^N y_i y_i'$, and W has the Wishart distribution with $N-2$ degrees of freedom, $W \sim W_p(\Sigma, N-2)$. Hence

$$Q = \frac{y_2' W^{-1} y_2}{1 + y_2' W^{-1} y_2} = 1 - \frac{1}{1 + y_2' W^{-1} y_2}.$$

Note that under H

$$\frac{N-2-p+1}{p} y_2' W^{-1} y_2 \sim F_{p, N-p-1}$$

where $F_{m,n}$ denotes F -distribution with p and $N-p-1$ d.f.

Thus, the hypothesis H is rejected if

$$\frac{N-p-1}{p} \frac{Q}{1-Q} = \frac{N-p-1}{p} y_2' W^{-1} y_2 > F_{p, N-p-1, \alpha},$$

where $F_{p, N-p-1, \alpha}$ is the upper $\alpha\%$ point of the F distribution with p and $N-p-1$ d.f. Equivalently, H is rejected if $Q \geq C_\alpha$ where $C_\alpha = 1 - [1 + p(N-p-1)^{-1} F_{p, N-p-1, \alpha}]^{-1}$.

4. ILLINOIS TRAFFIC DATA

Consider the Illinois Traffic Data given in Table 2. The table gives the annual data from 1962 to 1971 on hundreds of traffic deaths, thousands of traffic injuries, thousands of accidents, and the number of deaths per hundred million vehicle miles in the State of Illinois.

It was felt that the increase in traffic fatalities and accidents each year would tend to be constant. A *change* in the mean increase each year would then be attributed to external influences such as new regulations and safety standards. For example variables of interest would be the increase in traffic deaths or injuries from 1962 to 1963. Our interest in the increase in traffic deaths and injuries means that we must look at the new set of variables formed by taking differences.

TABLE 2: *Illinois traffic data from 1962 to 1971.*

	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Deaths (10^2)	18.90	20.28	22.07	22.56	25.22	24.93	24.99	25.33	23.46	24.00
Injuries (10^3)	112.31	119.89	134.16	145.54	149.14	149.51	148.73	157.45	159.88	148.83
Accidents (10^3)	252.02	258.68	281.16	324.07	329.42	337.56	351.07	405.51	409.17	393.57
Deaths per 10^8 vehicle miles	4.9	5.1	5.2	5.1	5.3	5.1	4.9	4.7	4.2	4.2

Sen and Srivastava (1975c) considered the four types of traffic data separately and applied univariate results. Here we apply multivariate techniques. Let

$$x_i^{(j)} = y_{i+1}^{(j)} - y_i^{(j)} = \mu_i^{(j)} + \varepsilon_i^{(j)}, \quad i = 1, 2, \dots, N,$$

$$j = 1, 2, 3, 4,$$

where $y_i^{(j)}$ is the j th type of traffic data for the i th year.

In vector notation, the above model becomes

$$\underline{x}_i = \underline{\mu}_i + \underline{\varepsilon}_i, \quad i = 1, 2, \dots, N.$$

We assume that $\underline{\varepsilon}_i$'s are independently normally distributed with mean vector zero and covariance matrix Σ . When Σ is known or $\Sigma = \sigma^2 I$, σ^2 unknown, Sen and Srivastava (1973) proposed tests for testing the hypothesis (1) against the alternative (2). In this section we test this hypothesis using the statistic given in Section 3 when Σ is completely unknown. An estimate of the change point 'r' is obtained by the likelihood method of Section 2. Using the results of Section 2, we get the values of T_r obtained as 2.64, 3.92, 6.21, 6.44, 1.71, 0.66, 1.55, 2.20. A change in mean was estimated to occur after the 4th interval. That is, there was a shift in the annual increase of traffic casualties after 1966. The value of the test statistic Q is 0.8750. With $C_{.05} = [1 - \{1 + 4\}(4)^{-1} \times 6.39\}^{-1}] = 0.865$, we

reject H and claim that there was a significant shift in the rise of traffic casualties in Illinois with the shift estimated to have occurred in 1966. From Table 1, the likelihood ratio tests reject H if $T_r \div (1 + T_r) > .94$. However

$$T_r \div (1 + T_r) = 6.44/7.44 = .85. \text{ Hence } H \text{ is not rejected.}$$

The discrepancy in the two results is attributed to the fact that there may be several changes in the mean and even if there is one change, the change occurs near the midpoint where the Bayes procedures has been shown in the univariate cases to be superior.

REFERENCES

- Bhattacharya, G. K. and Johnson, R. A. (1968). Nonparametric tests for shift at an unknown point *Annals of Mathematical Statistics*, 39, 1731-1743.
- Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics*, 35, 990-1018.
- Gardner, L. A. (1969). On detecting changes in the mean of normal variables *Annals of Mathematical Statistics*, 40, 116-126.
- Illinois Department of Transportation. (1972). Illinois Motor Vehicle Accident Facts, Springfield, Illinois.
- Page, E. S. (1955). A test for a change in parameter occurring at an unknown point. *Biometrika*, 42, 523-526.
- Sen, A. K. (1971). Tests for change in mean and a sequential ranking procedure. Ph.D. thesis, University of Toronto.
- Sen, A. K. and Srivastava, M. S. (1973). On multivariate tests for detecting change in mean. *Sankhyā*, A, 35, 173-185.
- Sen, A. K. and Srivastava, M. S. (1975a). On tests for detecting change in mean. *Annals of Statistics*, 3, 98-108.
- Sen, A. K. and Srivastava, M. S. (1975b). On tests for detecting change in mean when variance is unknown. *Annals of the Institute of Statistical Mathematics*, 479-486.
- Sen, A. K. and Srivastava, M. S. (1975c). Some one-sided tests for change in level. *Technometrics*, 17, 61-64.
- Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North Holland Publishing Co., New York.

[Received June 1980. Revised October 1980]

A TWO-DIMENSIONAL T-DISTRIBUTION AND A NEW TEST WITH FLEXIBLE TYPE I ERROR CONTROL

G. LANDENNA and D. MARASINI

Instituto di Scienze Statistiche e
Matematiche "Marcello Boldrini"
Universita degli Studi di Milano

SUMMARY. The usual t-test for a null hypothesis $H_0 = H_{01} \wedge H_{02}$, where $H_{0j}: \mu_j = \mu_j^*$, $j=1,2$, concerning the means of two normal populations with equal but unknown variance, does not account for the possibility that one of the component hypotheses may be principal, i.e., scientifically more relevant or important than the other. In this we use some properties of a bivariate t-distribution to construct a test which permits asymmetric treatment of the two component hypotheses. The proposed test is shown to be unbiased and consistent.

KEY WORDS. hypothesis testing, consistent tests, unbiased tests, t-tests, t-distribution.

1. INTRODUCTION AND SUMMARY

Consider the problem of testing a null hypothesis $H_0 = H_{01} \wedge H_{02}$, $H_{0j}: \mu_j = \mu_j^*$, $j=1,2$, where μ_1, μ_2 are the means of two normal populations with equal but unknown variance σ^2 . The traditional procedure for H_0 is a t-test based on the premise that the two component hypotheses H_{0j} are equally important, and permits control of only the overall type I error with respect to H_0 . However, it is not uncommon to have practical situations where one of the components is considerably more important than the other and it is imperative to adjust the test procedure for this. In this paper we discuss some properties of a bivariate t-

distribution and use these to construct a test which offers flexibility in the type I error control with respect to both the overall null hypothesis H_0 and its components H_{0j} . Specifically, we assume that H_{01} is the *principal* component, i.e., of primary interest, and propose a test for which

- i) the probability of the type I error with respect to H_0 equals a specified value α , $0 < \alpha < 1$;
- ii) the probability of the type I error with respect to H_{01} equals another specified value α' , $0 < \alpha' < \alpha < 1$.

The type I error control in this proposed test is analogous to that in the well-known step-down procedures in multivariate analysis as reviewed in Mudholkar and Subbaiah (1980). See also a study of a modification of such a procedure by Mudholkar and Subbaiah (1981).

In Section 2 we discuss some properties of a bivariate t -distribution. The new test with a flexible type I error control is presented in Section 3. Also in this final section, we show that the new test is unbiased and consistent, in the sense that the power function converges to unity as either the sample sizes or noncentrality parameters tend to infinity.

2. A BIVARIATE t -DISTRIBUTION

Let m_1 and m_2 denote the means of random samples of sizes n_1 and n_2 from two normal populations with a common variance σ^2 and means μ_1 and μ_2 respectively. Let s^2 denote the pooled estimate of σ^2 based upon the two samples. That is, let

$$s^2 = \frac{2}{\sum_{i=1}^2} s_i^2 / g,$$

where $g = \sum_{i=1}^2 g_i$, $g_i = n_i - 1$ and $s_i^2 = \Sigma(x_i - m_i)^2 / g_i$. Then the joint p.d.f. of m_1, m_2 and s^2 is

$$\phi(m_1, m_2, s^2) = \frac{(m_1, m_2)^{\frac{1}{2}} \frac{g}{2} (s^2)^{\frac{g}{2}-1} e^{-\frac{gs^2}{2\sigma^2} \left[1 + \left(\frac{m_1 - \mu_1}{\sqrt{g} s / \sqrt{m_1}} \right)^2 + \left(\frac{m_2 - \mu_2}{\sqrt{g} s / \sqrt{m_2}} \right)^2 \right]}}{\pi \Gamma\left(\frac{g}{2}\right) g^{\frac{g}{2}+1} (\sigma^2)^{\frac{g}{2}+1}} \quad (1)$$

Observing that the ratios

$$t_i = \frac{m_i - \mu_i}{s / \sqrt{n_i}}, \quad i=1, 2, \quad (2)$$

appearing in (1) are distributed as usual Student's t 's, and integrating with respect to s^2 we obtain the joint p.d.f. of t_1 and t_2 as

$$\psi(t_1, t_2) = \frac{\Gamma\left(\frac{g+2}{2}\right)}{\Gamma\left(\frac{g}{2}\right) \pi g} \left(1 + \frac{t_1^2}{g} + \frac{t_2^2}{g} \right)^{-\frac{g+2}{2}} \quad (3)$$

This is the p.d.f. of a particular case of the well-known multivariate t -distribution, e.g., see Johnson and Kotz (1972). The marginal p.d.f.'s of this bivariate t -distribution with g degrees of freedom are the familiar Student's t -distributions given by

$$\psi_i(t_i) = \frac{\Gamma\left(\frac{g+1}{2}\right)}{\Gamma\left(\frac{g}{2}\right) \sqrt{\pi g}} \left(1 + \frac{t_i^2}{g} \right)^{-\frac{g+1}{2}}, \quad i=1, 2. \quad (4)$$

On the other hand, the conditional distributions are not Student's t -distributions unless the conditioning r.v. takes values ± 1 . For example, the conditional p.d.f. of t_2 given t_1 is

$$\eta_2(t_2|t_1) = \frac{\psi(t_1, t_2)}{\psi_1(t_1)} = \frac{\Gamma\left(\frac{g+2}{2}\right)}{\Gamma\left(\frac{g+1}{2}\right) \sqrt{\pi(g+t_1^2)}} \left(1 + \frac{t_2^2}{g+t_1^2}\right)^{-\frac{g+2}{2}} \quad (5)$$

which, when $t_1 = \pm 1$ reduces to the p.d.f. of the Student's r.v. with $(g+1)$ degrees of freedom.

Lemma 1. Let (T_1, T_2) be a two-dimensional Student's r.v. with p.d.f. (3). If ϵ is any positive constant, then:

$$P_\epsilon = \int_{-\infty}^{\epsilon(g+t_1^2)^{\frac{1}{2}}} \eta_2(t_2|t_1) dt_2 = \int_{-\infty}^{\epsilon(g+1)^{\frac{1}{2}}} \eta_2(t_2|\pm 1) dt_2. \quad (6)$$

Proof. We have:

$$\begin{aligned} P_\epsilon &= \int_{-\infty}^{\epsilon(g+t_1^2)^{\frac{1}{2}}} \eta_2(t_2|t_1) dt_2 \\ &= \frac{\Gamma\left(\frac{g+2}{2}\right)}{\Gamma\left(\frac{g+1}{2}\right) \sqrt{\pi(g+t_1^2)}} \int_{-\infty}^{\epsilon(g+t_1^2)^{\frac{1}{2}}} \left(1 + \frac{t_2^2}{g+t_1^2}\right)^{-\frac{g+2}{2}} dt_2. \end{aligned}$$

Letting $t_2^2 = (g+t_1^2) y(1-y)^{-1}$, we obtain:

$$P_\epsilon = \frac{1}{2} + \frac{\Gamma\left(\frac{g+2}{2}\right)}{2 \Gamma\left(\frac{g+1}{2}\right) \sqrt{\pi}} \int_0^a y^{\frac{1}{2}-1} (1-y)^{\frac{g+1}{2}-1} dy, \dots \quad (7)$$

where

$$a = \varepsilon^2(1+\varepsilon^2)^{-1} \quad (8)$$

From the fact that both (7) and (8) are independent of the particular value assumed by the conditioning r.v.'s we have (6).

Now recalling that the marginals of (T_1, T_2) are Student's r.v. with g degrees of freedom from Lemma 1 we derive the following:

Lemma 2. Let (T_1, T_2) be a two-dimensional Student's r.v. with p.d.f. (3). If α' and α are any real constants $(0 < \alpha' < \alpha < 1)$ and t_1^* is the solution of

$$\int_{-\infty}^{t_1^*} \psi_1(t_1) dt_1 = 1 - \alpha', \quad (9)$$

then there exists a positive value ε^* such that:

$$\iint_R \psi(t_1, t_2) dt_1 dt_2 = 1 - \alpha. \quad (10)$$

Here R is the region of the (t_1, t_2) -plane defined as follows:

$$R: \begin{cases} -\infty < t_1 < t_1^* \\ -\infty < t_2 < \varepsilon^* (g + t_1^2)^{\frac{1}{2}} \end{cases} \quad (11)$$

Proof. We can write:

$$\begin{aligned}
 (1-\alpha) &= \iint_{\mathbb{R}} \psi(t_1, t_2) dt_1 dt_2 = \int_{-\infty}^{t_1^*} \int_{-\infty}^{\varepsilon^*(g+t_1)^{\frac{1}{2}}} \psi(t_1, t_2) dt_1 dt_2 \\
 &= \int_{-\infty}^{t_1^*} \psi_1(t_1) \int_{-\infty}^{\varepsilon^*(g+t_1)^{\frac{1}{2}}} \eta_2(t_2 | t_1) dt_2 dt_1.
 \end{aligned}$$

But by Lemma 1 we have

$$(1-\alpha) = \int_{-\infty}^{t_1^*} \psi_1(t_1) \int_{-\infty}^{\varepsilon^*(g+1)^{\frac{1}{2}}} \eta_2(t_2 | 1) dt_2 dt_1.$$

Hence from (9) we get

$$(1-\alpha) = (1-\alpha') \int_{-\infty}^{\varepsilon^*(g+1)^{\frac{1}{2}}} \eta_2(t_2 | 1) dt_2.$$

That is,

$$\int_{-\infty}^{\varepsilon^*(g+1)^{\frac{1}{2}}} \eta_2(t_2 | 1) dt_2 = \frac{1-\alpha}{1-\alpha'}. \quad (12)$$

If t_2^* is the solution of (9) we have:

$$t_2^* = \varepsilon^*(g+1)^{\frac{1}{2}}.$$

Hence

$$\varepsilon^* = t_2^*(g+1)^{-\frac{1}{2}}.$$

The region (11) may now be expressed as:

$$R: \begin{cases} -\infty < t_1 < t_1^* \\ -\infty < t_2 < t_2^{**} = t_2^* (g+1)^{-\frac{1}{2}} (g+t_1^2)^{\frac{1}{2}}, \end{cases} \quad (13)$$

where t_1^* and t_2^* are the critical constants corresponding to levels of significance α' and $(1-\alpha)/(1-\alpha')$ which we obtain from the one-dimensional Student's tables.

3. THE NEW TEST

Consider two normal populations with equal but unknown variance σ^2 and unknown means μ_1 and μ_2 . We want to test the null hypothesis

$$H_0 = H_{01} \wedge H_{02}, \text{ with } H_{0j}: \mu_j = \mu_j^*,$$

$j=1,2$, assuming H_{01} as principal. Let α be the level of significance for H_0 , and α' that for H_{01} , $0 < \alpha' < \alpha < 1$. Suppose we have random samples of sizes n_1 and n_2 from these populations and we have obtained

$$t_1 = \frac{m_1 - \mu_1^*}{s/\sqrt{n_1}} \quad \text{and} \quad t_2 = \frac{m_2 - \mu_2^*}{s/\sqrt{n_2}}$$

where m_1 and m_2 are the sample means and s^2 is the pooled estimate of common variance σ^2 . Then (t_1, t_2) has the sampling distribution of a two-dimensional Student's r.v. (T_1, T_2) with $g = n_1 + n_2 - 2$ degrees of freedom. It follows that for testing hypothesis $H_0: H_{01} \wedge H_{02}$ we consider the region (13) and we

accept H_0 if $(t_1, t_2) \in R$, otherwise we reject it. This procedure for R ensures the control of the type I error with respect to both H_0 and H_{01} .

Now consider the alternative hypotheses:

$$\bar{H}_{01} \wedge \bar{H}_{02}, \quad \bar{H}_{01} \wedge H_{02}, \quad H_{01} \wedge \bar{H}_{02},$$

where: $\bar{H}_1: \mu_1 < \mu_1^*$, $\bar{H}_2: \mu_2 < \mu_2^*$. We now show the unbiasedness of this test and the convergence to unity of the power function with respect to the above mentioned alternative hypotheses for large values of the corresponding non-centrality parameters λ_1, λ_2 . Towards this end we write:

$$\bar{H}_1 \wedge \bar{H}_2: \lambda_1 > 0, \lambda_2 > 0,$$

$$\bar{H}_1 \wedge H_2: \lambda_1 > 0, \lambda_2 = 0,$$

$$H_1 \wedge \bar{H}_2: \lambda_1 = 0, \lambda_2 > 0.$$

In doing so, we note that a one-dimensional Student's non-central T can always be written in the form:

$$\tilde{T} = \frac{M - \mu + \delta}{S/\sqrt{n}} = \left[\frac{M - \mu}{\sigma/\sqrt{n}} + \frac{\sqrt{n}\delta}{\sigma} \right] \left(\frac{\chi^2}{g} \right)^{-\frac{1}{2}} = \frac{Z}{(\chi^2/g)^{1/2}} + \frac{\lambda}{(\chi^2/g)^{1/2}},$$

where M and S are the sample mean and the sample standard deviation respectively, Z is a r.v. having normal distribution in the standardized form, χ^2 has g degrees of freedom, and $\lambda = \sqrt{n}\delta/\sigma$ is the non-centrality parameter, $\delta = (\mu - \mu_0)$.

Therefore, considering the two-dimensional Student's non-central r.v. $(\tilde{T}_1, \tilde{T}_2)$ with non-centrality parameters:

$$\lambda_1 = \frac{\sqrt{n_1} \delta_1}{\sigma}, \quad \lambda_2 = \frac{\sqrt{n_2} \delta_2}{\sigma},$$

where: $\delta_1 = \mu_1 - \mu_1^*$ and $\delta_2 = \mu_2 - \mu_2^*$, we can write the probability of committing a type I error as follows:

$$\begin{aligned} \beta(\lambda_1, \lambda_2) &= \Pr\{\tilde{T}_1(\lambda_1) < t_1^*, \tilde{T}_2(\lambda_2) < t_2^{**}\} \\ &= \Pr\left\{\frac{Z_1}{(\chi^2/g)^{1/2}} + \frac{\lambda_1}{(\chi^2/g)^{1/2}} < t_1^*, \frac{Z_2}{(\chi^2/g)^{1/2}} + \frac{\lambda_2}{(\chi^2/g)^{1/2}} < t_2^{**}\right\} \\ &= \Pr\left\{Z_1 - t_1^* \left(\frac{\chi^2}{g}\right)^{\frac{1}{2}} < -\lambda_1, Z_2 - t_2^{**} \left(\frac{\chi^2}{g}\right)^{\frac{1}{2}} < -\lambda_2\right\}, \quad (14) \end{aligned}$$

where t_1^* and t_2^{**} are the critical values according to the region (13). For large values of λ_1 or λ_2 in (14) we obtain:

$$\lim_{\lambda_i \rightarrow \infty} \beta(\lambda_1, \lambda_2) = 0 \quad (i=1,2)$$

so that the power function, $\theta(\lambda_1, \lambda_2) = 1 - \beta(\lambda_1, \lambda_2)$, converges to unity.

From (14) it also follows that $\beta(\lambda_1, \lambda_2)$ is always a decreasing function of λ_1 and λ_2 . From this fact and observing that for $\lambda_1 = 0$, $\lambda_2 = 0$, (15) is of the form:

$$\beta(0,0) = \Pr\left\{Z_1 - t_1^* \left(\frac{\chi^2}{g}\right)^{1/2} < 0, Z_2 - t_2^{**} \left(\frac{\chi^2}{g}\right)^{1/2} < 0\right\}$$

$$= \Pr \left\{ \frac{Z_1}{(\chi^2/g)^{1/2}} < t_1^*, \frac{Z_2}{(\chi^2/g)^{1/2}} < t_2^{**} \right\}$$

$$= \Pr \{ \tilde{T}_1(0) < t_1^*, \tilde{T}_2(0) < t_2^{**} \} = 1 - \alpha,$$

it follows that the power function $\theta(\lambda_1, \lambda_2)$, for every pair (λ_1, λ_2) , with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $(\lambda_1 + \lambda_2) \neq 0$, is greater than α and therefore the test is unbiased with respect to $\bar{H}_1 \wedge \bar{H}_2$.

In an analogous manner it can be shown that the above properties hold with respect to alternatives $\bar{H}_1 \wedge H_2$ and $H_1 \wedge \bar{H}_2$. The test can be easily modified to test $H_0 = H_{01} \wedge H_{02}$ against two-sided alternatives of the form $\bar{H}_1: \mu_1 \neq \mu_1^*$, $\bar{H}_2: \mu_2 \neq \mu_2^*$.

Acknowledgement. We wish to express our gratitude to G. S. Mudholkar for his assistance in revising an earlier draft of this article.

REFERENCES

- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Mudholkar, G. S. and Subbaiah, P. (1980). A review of step-down procedures for multivariate analysis of variance. In *Multivariate Statistical Analysis*, R. P. Gupta, ed. North Holland, New York.
- Mudholkar, G. S. and Subbaiah, P. (1981). Complete independence in the multivariate normal distribution. In *Statistical Distributions in Scientific Work*, C. Taillie, G. P. Patil, and B. Baldessari, eds. Reidel, Dordrecht-Holland. Vol. 5, pp. 157-168.

[Received July 1980. Revised December 1980]

TESTING OUTLIERS IN MULTIVARIATE DATA

M. L. TIKU and M. SINGH

Department of Mathematical Sciences
McMaster University
Hamilton, CANADA L8S 4K1

SUMMARY. Given n random observations on a p -dimensional random vector x , the problem is to test whether a specified number (usually small) of suspected observations are outliers (too discordant as compared to the bulk of observations). As a generalization of Tiku's (1975, 1977) univariate statistic, we propose a statistic g for testing a specified number of outliers in multivariate data; g is the ratio of the product of robust estimators (Tiku, 1980) to the product of ordinary estimators of the scale parameters. For the multivariate normal, g is shown to be considerably more powerful than the prominent statistic R (restricted to the multivariate normal) due to Wilks (1963) under location shifts (model A; Barnett and Lewis, 1978) although slightly less powerful under scale changes (model B; Barnett and Lewis). Like R , g is not sensitive to changes in correlations (orientation). The statistic g can be used (under models A or B) for testing outliers in samples from any multivariate distribution whose marginal distributions are of the type $(1/\sigma)f((x-\mu)/\sigma)$.

KEY WORDS. Multivariate outliers, censored samples, robust estimators, modified maximum likelihood estimators.

1. FORMULATION THROUGH MARGINAL SAMPLES

Barnett and Lewis (1978, pp. 220, 221) remark: We should not underestimate the role to be played by the marginal samples (that is, the univariate samples of each component value in the multivariate data) in the identification of outliers. Firstly, we know what we mean by a marginal sample. Secondly, we have

facilities for testing the discordancy of such univariate outliers for a range of different basic models (and we can adopt models to explain the outliers). And thirdly, perhaps most important, it is quite plausible to expect outliers to be exhibited within specific components of the multivariate observations.

In the spirit of this remark, let $(x_{1i}, x_{2i}, \dots, x_{pi})$, $i = 1, 2, \dots, n$, be a random sample from a multivariate distribution $f(x_1, x_2, \dots, x_p)$; $E(x_k) = \mu_k$, $V(x_k) = \sigma_k^2$ and $\text{Cov}(x_k, x_\ell) = \rho_{k\ell} \sigma_k \sigma_\ell$, $\rho_{kk} = 1$. In the available n observations

$$x_{k1}, x_{k2}, \dots, x_{kn} \quad (1)$$

on the k th component x_k , however, $r_{k1} (> 0)$ smallest and $r_{k2} (> 0)$ largest observations appear to be too small and too large (potential outliers), respectively, as compared with the bulk of observations in (1); $k = 1, 2, \dots, p$. The problem is to test whether these suspected observations are in fact outliers.

For the univariate case ($p = 1$), a number of solutions to this problem are known; see Tietjen and Moore (1972) and Tiku (1975, 1977) and the references cited in these papers. For the multivariate case ($p \geq 2$), this problem is more complex and, therefore, the solutions are relatively sparse; see, however, Wilks (1963), Cox (1968), Healy (1968), Gnanadesikan and Ketternring (1972), Barnett and Lewis (1978, Chapter 6) and Barnett (1979).

2. THE TEST STATISTIC

$$\text{Let } x_{r_{k1}+1}, x_{r_{k1}+2}, \dots, x_{n-r_{k2}} \quad (2)$$

be the Type II censored sample, obtained by arranging the observations in the marginal sample (1) in ascending order of magnitude and censoring the r_{k1} smallest and the r_{k2} largest observations; $k = 1, 2, \dots, p$. In the bivariate case it will be convenient to write $r = (r_{11}, r_{12}; r_{21}, r_{22})$. Let $\hat{\sigma}_k$ be the robust estimator (modified maximum likelihood estimator; Tiku, 1967, 1980) based on the censored sample (2) and s_k be the ordinary estimator based on the complete sample (1). In fact, s_k is the estimator $\hat{\sigma}_k$ with $r_{k1} = r_{k2} = 0$; see the Appendix. Pretend that the correlations $\rho_{k\ell}$ are known. A robust estimator of the "internal scatter" of the multivariate sample is then given by $|(\rho_{k\ell} \hat{\sigma}_k \hat{\sigma}_\ell)|$ and the corresponding ordinary estimator is

given by $|(\rho_{k\ell} s_k s_\ell)|$. Consider the ratio

$$g = \sqrt{\{ |(\rho_{k\ell} \hat{\sigma}_k \hat{\sigma}_\ell)| / |(\rho_{k\ell} s_k s_\ell)| \}} = \prod_{k=1}^p (\hat{\sigma}_k / s_k), \quad (3)$$

since $|(\rho_{k\ell} \hat{\sigma}_k \hat{\sigma}_\ell)| = (\hat{\sigma}_1^2 \hat{\sigma}_2^2 \cdots \hat{\sigma}_p^2) |(\rho_{k\ell})|$ and $|(\rho_{k\ell} s_k s_\ell)| = (s_1^2 s_2^2 \cdots s_p^2) |(\rho_{k\ell})|$. We propose g as a test statistic for testing suspected outliers (see also Tiku, 1975, 1977); small values of g indicate the presence of outliers among the n data points \tilde{x}_i . The statistic g is a natural generalization of Tiku's (1975, 1977) univariate statistic $\hat{\sigma}/s$. The explicit expressions of $\hat{\sigma}_k$ and s_k for the normal, exponential and uniform populations are given in the Appendix. Note that g is location and scale invariant. However, g takes no explicit account of correlations (orientation) of the data. A statistic g^* which is similar to g but takes explicit account of the correlations is developed in Section 4.

The null (no outliers exist) distribution of $h(\hat{\sigma}_k/s_k)$, h being a constant, is exactly Beta for the exponential and uniform populations and approximately Beta for the normal population; see the Appendix. Note that the p ratios $\hat{\sigma}_k/s_k$, $k = 1, 2, \dots, p$, are uncorrelated (independent for the multivariate normal and certain types of exponential and uniform populations; see the Appendix) if all the correlations $\rho_{k\ell}$ are zero, but not so if some $\rho_{k\ell}$ are non-zero. Evidently then the null distribution of g depends on the correlations $\rho_{k\ell}$. Luckily, however, this dependence is not 'strong', that is, the mean and variance and the percentage points of g do not change much with changes in $\rho_{k\ell}$, at any rate for the bivariate normal. For the bivariate normal, the simulated (based on 10,000 Monte Carlo runs) values of the mean and variance and the lower 5 and 10 percent points of g are given in Table 1. It is clear that these values do not change much with changes in ρ ; see also Table 3. The same of course is true for the bivariate exponential and uniform populations but we omit details for conciseness; see, however, Table 3. The percentage points of g for $\rho = 0$ can, therefore, be used also for non-zero values of ρ . The equations for calculating the approximate percentage points of g are given in the Appendix.

Two most important models for generating a single outlier are the following (Barnett and Lewis, 1978, p. 210; Barnett, 1978):

TABLE 1: Simulated values of the mean and variance and the lower 5 and 10 percent points of the statistic g for the bivariate normal.

	n=10					n=20				
	ρ					ρ				
	-.8	-.5	0	.5	.8	-.8	-.5	0	.5	.8
$r = (0, 1; 0, 1)$										
Mean	.984	.985	.984	.983	.985	.997	.997	.996	.995	.995
Var.	.018	.019	.020	.021	.024	.0041	.0044	.0044	.0047	.0056
5%	.728	.717	.712	.705	.678	.875	.869	.866	.862	.845
10%	.795	.788	.787	.787	.773	.909	.908	.905	.903	.897
$r = (0, 1; 1, 0)$										
Mean	.988	.986	.984	.982	.981	.998	.997	.997	.997	.997
Var.	.023	.019	.019	.019	.019	.0053	.0045	.0044	.0043	.0042
5%	.688	.716	.716	.710	.717	.853	.865	.868	.871	.875
10%	.779	.788	.788	.790	.794	.903	.908	.906	.909	.910
$r = (1, 1; 1, 1)$										
Mean	.972	.967	.965	.964	.968	.995	.993	.992	.991	.991
Var.	.052	.046	.045	.046	.052	.011	.0095	.0093	.0097	.011
5%	.548	.582	.597	.580	.548	.794	.813	.819	.810	.793
10%	.648	.672	.676	.670	.651	.847	.862	.860	.857	.844

Model A: $E(\tilde{x}_1) = \tilde{\mu} + \tilde{a}$ (some i)

$$E(\tilde{x}_j) = \tilde{\mu} \quad (j \neq 1)$$

with variance-covariance matrix $V(\tilde{x}_j) = V$

$$(j = 1, 2, \dots, n).$$

Model B: $V(\tilde{x}_1) = bV$ (some i) ($b > 1$)

$$V(\tilde{x}_j) = V \quad (j \neq 1)$$

with mean vector $E(\tilde{x}_j) = \tilde{\mu} \quad (j = 1, 2, \dots, n).$

The mean and variance of all the components x_k can without any loss of generality be assumed to be 0 and 1, respectively. Generalization of the above models to more than one outlier is straightforward.

Model A tends to generate a single 'inflated' (too small or too large) order statistic only on one side (depending on the sign of the k th component a_k of \tilde{a}) of the marginal sample (1).

Model B tends to generate a single 'inflated' order statistic on either side of the marginal sample (1); $E(x_k) = \mu_k = 0$. This governs the choice of the values of r_{k1} and r_{k2} ($k = 1, 2, \dots, p$), the number of smallest and largest observations, respectively, censored in the marginal sample (1), for calculating g . Under model A, for example, if a is greater than 0 (in fact, substantially greater than 0) then $r_{k1} = 0$ and $r_{k2} = 1$, $k = 1, 2, \dots, p$; (see Table 2). Under model B, $r_{k1} = r_{k2} = 1$ for all k . Another model which typically generates a single 'inflated' order statistic is Tiku's (1975, 1977) outlier model (labeled slippage model of Barnett, 1978, p. 249) which adds a constant to the largest order statistic or subtracts a constant from the smallest order statistic of a random sample; under this model the statistic g , like the univariate statistic $\hat{\sigma}/s$, is particularly powerful; see also Tiku (1975, 1977) and Hawkins (1977). In the rest of this paper, therefore, we study the performance (power properties) of g only under models A and B; power defined to be the percentage of the time a single (in general, a specified number $r \geq 1$) outlier is detected which is perhaps the most useful of the five measures (David and Paulson, 1965) of the performance of an outlier test; see Hawkins (1977, p. 436).

3. POWER COMPARISONS

For testing a single outlier in the multivariate normal sample, the prominent statistic due to Wilks (1963) is given by

$$R = \min \{ |\hat{\rho}_{kl}^* s_k^* s_l^*| / |(\hat{\rho}_{kl} s_k s_l)| \} \quad (4)$$

where the numerator is the maximum likelihood estimator of the 'internal scatter' obtained by deleting one (out of n) observation $(x_{1i}, x_{2i}, \dots, x_{pi})$, $i = 1, 2, \dots, n$, at a time, and the denominator is the maximum likelihood estimator based on all the n observations; R is thus the minimum of n ratios (of the determinants of two $p \times p$ matrices). Note that R is location and scale invariant and the null distribution of R is invariant with respect to the correlations ρ_{kl} (Wilks, 1968). The exact null distribution of R is not tractable (Wilks, 1963; Barnett, 1979); see, however, Siotani (1959) who tabulates the approximate percentage points of a similar statistic. Note that for testing a specified number $r \geq 2$ of outliers, the statistic R is the minimum of a large number $(n!/r!(n-r)!)$ of ratios like (4) and the computation of R is therefore problematic. The computation of g does not pose any problem, for testing any specified number of outliers.

TABLE 2: Simulated values of the power of g , for testing a single outlier in bivariate normal samples; $\sigma_1 = \sigma_2 = 1$; $a = (a_1, a_2)$.

Model A: (x_{11}, x_{21}) replaced by $(x_{11} + a_1\sigma_1, x_{21} + a_2\sigma_2)$																
$r = (0, 1; 0, 1)$																
$a = (3, 3)$				$a = (3, 1)$				$a = (1, 1)$								
$\rho = 0$		$\rho = .5$		$\rho = 0$		$\rho = .5$		$\rho = 0$		$\rho = .5$						
R	g^*	R	g^*	R	g^{**}	R	g^{**}	R	g^{***}	R	g^{***}	R	g^{***}			
$n = 10$																
5%	.38	.53	.24	.50	.20	.28	.19	.28	.06	.06	.06	.07				
10%	.53	.67	.36	.63	.31	.40	.29	.40	.11	.12	.11	.12				
$n = 20$																
5%	.56	.57	.35	.53	.29	.33	.25	.33	.07	.07	.06	.07				
10%	.68	.69	.48	.64	.40	.46	.36	.45	.12	.13	.12	.13				
$r = (0, 1; 1, 0)$																
$a = (3, -3)$				$a = (3, -1)$				$a = (1, -1)$								
$\rho = 0$		$\rho = -.5$		$\rho = 0$		$\rho = -.5$		$\rho = 0$		$\rho = -.5$						
R	g^*	R	g^*	R	g^{**}	R	g^{**}	R	g^{***}	R	g^{***}	R	g^{***}			
$n = 10$																
5%	.38	.55	.24	.53	.19	.30	.18	.30	.05	.06	.05	.07				
10%	.53	.67	.37	.64	.31	.42	.30	.42	.10	.12	.10	.13				
$n = 20$																
5%	.55	.58	.32	.53	.26	.33	.25	.31	.06	.06	.06	.06				
10%	.67	.68	.44	.63	.37	.43	.34	.42	.12	.12	.11	.12				
Model B: (x_{11}, x_{21}) replaced by (bx_{11}, bx_{21})																
$r = (1, 1; 1, 1)$																
$\rho=0$				$\rho=.5$				$\rho=0$				$\rho=.5$				
$b=2$		$b=8$		$b=2$		$b=8$		$b=2$		$b=8$		$b=2$		$b=8$		
R	g	R	g	R	g	R	g	R	g	R	g	R	g	R	g	
$n=10$								$n=20$								
5%	.12	.12	.80	.80	.12	.12	.80	.77	.17	.16	.86	.85	.17	.13	.86	.83
10%	.20	.20	.85	.84	.20	.20	.85	.82	.25	.22	.89	.88	.25	.21	.89	.86

For $r_{11}=0$, $r_{12}=1$ and $r_{21}=r_{22}=0$, the values of the power (for 5 and 10 percent significance levels, respectively) of g are as follows:

- * 0.39 and 0.53 for $n=10$, and 0.40 and 0.52 for $n=20$; for all values of ρ .
- ** 0.39 and 0.53 for $n=10$, and 0.40 and 0.52 for $n=20$; for all values of ρ .
- *** 0.07 and 0.14 for $n=10$, and 0.06 and 0.12 for $n=20$; for all values of ρ .

For bivariate normal population, the simulated (based on 2000 runs) values of the power of the statistics R and g for testing a single outlier, generated under models A and B, are given in Table 2, for $\rho = -.5, 0$ and $.5$. The values for $\rho = -.8$ and $.8$ were similar (rather more favorable to g) and are not reproduced for brevity. It is clear that R and g are ineffective if both the components of the suspected observation are inliers (rather than outliers) in the respective marginal samples, i.e., a_1 and a_2 are small, irrespective of the values of ρ .

On the other hand, R and g are effective if either of the two components of the suspected observation are 'inflated' extreme order statistics (potential outliers) in the respective marginal samples, i.e., $|a_1|$ or $|a_2|$ are large; see also

Examples 1 to 3. Under model A, g is more powerful than R . Under model B, however, g is slightly less powerful than R and this is to be expected since R is 'optimal' under model B; see Barnett and Lewis (1978, p. 219). We would expect similar results for p -variate normal population ($p > 2$) and for the detection of more than one outlier but this needs further investigation.

Another important model is that of orientation (correlation changes). For example, a single outlier in the bivariate situation could be generated as follows:

Model C: corr. coeff. $(x_{1i}, x_{2i}) = \rho$ (some i)

corr. coeff. $(x_{1j}, x_{2j}) = \rho_0 \neq \rho$ ($j \neq i$)

$E(x_{li}) = E(x_{lj})$ and $V(x_{li}) = V(x_{lj})$, $l = 1, 2$.

Simulations reveal that both R and g (irrespective of the choices of r_{k1} and r_{k2}) are ineffective for detecting such outliers; for example for $\rho = 0.8$ and $\rho_0 = 0.0$, the value of the power of R and g [$r = (0, 1; 0, 1)$] are as follows:

	$n = 10$		$n = 20$	
	R	g	R	g
5%	.06	.05	.05	.05
10%	.11	.10	.11	.10

The procedures proposed by Gnanadesikan and Kettenring (1972) may be useful here; see also Healy (1968). Note, however, that no statistic can be expected to detect all types of outliers in multivariate data (Barnett and Lewis, 1978, p. 220).

Example 1. Consider the following 10 observations supposed to come from a bivariate normal distribution:

```

x1: .4930 .0280 1.618 -.7700 -.059
x2: -.1425 .2675 1.702 -1.1835 .012

x1: -.912 .378 -.2240 -.2210 -.3290
x2: -.974 .049 .0265 .2445 .0765

```

The third pair (1.618, 1.702) is suspect; since both of its components are the largest order statistics in the respective marginal samples (potential outliers), we choose $r_{11} = r_{21} = 0$ and $r_{12} = r_{22} = 1$ for calculating the statistic g . The value of g and the minimum value of R (which corresponds to the pair (1.618, 1.702)), and the 5% points are as follows:

	R	g
Calculated value	0.35	0.54
5% point	0.22	0.71 (for $\rho = 0$, Table 1).

The above data was in fact obtained by generating 10 random observations from a bivariate normal distribution ($\rho = 0.85$) and replacing a pair (x_{1i}, x_{2i}) , the third one, by $(x_{1i}+2, x_{2i}+2)$; the 5% point of g for $\rho = .85$ is in fact 0.68. The statistic g detects the outlier but not R , at 5% significance level.

Example 2. Consider the following 10 observations supposed to come from a bivariate normal:

```

x1: -.038 -.211 .604 -1.59 -.597
x2: 4.424 1.67 .669 .062 1.368

x1: .22 1.449 -.841 1.714 .932
x2: -.29 -.395 .786 -.573-1.003

```

The observation $(-.038, 4.424)$ is suspect; since the component $-.308$ is an inlier in the marginal x_1 -sample but the component 4.424 is a potential outlier (the largest order statistic) in the marginal x_2 -sample, we choose $r = (0, 0; 0, 1)$ for calculating g . The value of g and the minimum value of R (which corresponds to the pair $(-.038, 4.424)$), and the 5% points are as follows:

	R	g
Calculated value	0.24	0.68
5% point	0.22	0.79 ;

the percentage points of g for this situation are given by Tiku (1975, Table 1). The above observations were in fact obtained by generating 10 random observations from a bivariate normal distribution ($\rho = 0$) and replacing a pair (x_{1i}, x_{2i}) , the first one, by $(x_{1i}+0.5, x_{2i}+4)$; g detects the outlier but not R , at 5% significance level.

Example 3. Consider the following 10 observations supposed to come from a bivariate normal:

x_1 :	-0.277	-1.23	-0.909	.308	-.062
x_2 :	-.082	-.762	.747	1.749	-.121
x_1 :	-1.659	-.177	-.015	-.018	-.723
x_2 :	.523	.339	-.523	2.2024	-.997 .

It is difficult to tell which particular pair is suspect, if at all; perhaps $(-1.659, .523)$ or $(-.018, 2.2024)$. To calculate g we take $r = (1, 0; 0, 0)$ for the first pair and $r = (0, 0; 0, 1)$ for the second pair. The values of g and the minimum value of R (which corresponds to the pair $(-1.659, .523)$) are as follows:

	R	g
Calculated value	0.45	0.96 (for the pair $(-1.659, .523)$)
	0.63	1.05 (for the pair $(-.018, 2.2024)$)
5% point	0.22	0.79;

0.63 is the second smallest value of R . The percentage points of g are identical for the two pairs because of the symmetry of bivariate normal and the fact that the marginal distributions are also symmetric and normal. The above observations were obtained as in the earlier examples, with the seventh pair replaced by $(x_{1i}+1, x_{2i}+1)$; both R and g are ineffective here. Using g one could, of course, very easily test the significance of the two pairs $(-1.659, .523)$ and $(-.018, 2.2024)$ simultaneously, i.e., calculating g with $r = (1, 0; 0, 1)$. The resulting value of g , however, turns out to be not significant, as expected; both R and g miss the inlier $(-.177, .339)$.

There might be situations (we had hard time to generate one) when R will succeed in detecting an outlier but not g . However, it is clear from the above examples and the values given in Table 2 that g will succeed more often than R , under model A.

4. AN ALTERNATIVE STATISTIC

The statistic g was developed under the pretension that the correlation coefficients ρ_{kl} are known. For the multivariate normal population, it is possible to define a statistic, say g^* , without this pretension; g^* is the square root of the ratio of the determinants of 'robust' variance-covariance matrix $(\hat{\sigma}_{kl})$ and 'ordinary' variance-covariance matrix (s_{kl}) . The estimators $\hat{\sigma}_{kk}$ are the same as $\hat{\sigma}_k^2$ (eq. 3) and the estimators $\hat{\sigma}_{kl}$ are given by

$$\hat{\sigma}_{kl} = (\hat{\sigma}_k^2 + \hat{\sigma}_l^2 - \hat{\sigma}_D^2)/2$$

where $\hat{\sigma}_D^2$ is the robust estimator of the variance $\hat{\sigma}_D^2 (= \sigma_k^2 + \sigma_l^2 + 2\rho_{kl} \sigma_k \sigma_l)$ calculated from the n values

$$z_i = \begin{cases} x_{ki} - x_{li} & \text{if } s_{kl} > 0 \\ x_{ki} + x_{li} & \text{if } s_{kl} \leq 0, \quad i = 1, 2, \dots, n; \end{cases}$$

$s_{kl} = \sum_{i=1}^n (x_{ki} - \bar{x}_k)(x_{li} - \bar{x}_l)$. Remember to change the sign of $\hat{\sigma}_{kl}$ if $s_{kl} \leq 0$. Note that the sample z_1, z_2, \dots, z_n will hopefully have fewer outliers than either of the two samples $x_{k1}, x_{k2}, \dots, x_{kn}$ and $x_{l1}, x_{l2}, \dots, x_{ln}$; $\hat{\sigma}_D^2$ is calculated exactly the same way as $\hat{\sigma}_k^2$ (or $\hat{\sigma}_l^2$) with the ordered observations x_{ki} (or x_{li}) replaced by the ordered observations z_i , and r_{k1} and r_{k2} (or r_{l1} and r_{l2}), the number of smallest and largest observations censored, replaced by $r_1^* = \max(r_{k1}, r_{l1})$ and $r_2^* = \max(r_{k2}, r_{l2})$, respectively. The efficiency-properties of the variance-covariance matrix $(\hat{\sigma}_{kl})$ are discussed by Tiku and Singh (1980); suffice it to say here that in the presence of outliers $\hat{\sigma}_{kl}$ have considerably smaller mean square errors than s_{kl} . Note that for normal samples $\hat{\sigma}_{kl}$ is unbiased for large n , since $\hat{\sigma}_k^2$, $\hat{\sigma}_l^2$ and $\hat{\sigma}_D^2$ are all unbiased for large n ; see Tiku (1978, Lemma 2). The statistic g^* is thus given by

$$g^* = \sqrt{(|\hat{\sigma}_{kl}| / |s_{kl}|)};$$

it is possible that a value of $|(\hat{\sigma}_{k\ell})|$ less than zero might occur but such a value is replaced by zero. Small values of g^* indicate the presence of outliers. At the present time, we have no clue how to obtain the null (exact or approximate) distribution of g^* . However, we carried out a simulation study of the power-properties of g^* and the results are given below for the bi-variate normal population, $n = 20$:

Simulated values of the power (single outlier)									
Model A: $r = (0, 1; 0, 1)$									
(a_1, a_2)									
$\rho = 0$					$\rho = .5$				
Percent	Point	(3,3)	(3,1)	(1,1)	Percent	Point	(3,3)	(3,1)	(1,1)
5%	.87	.54	.33	.07	5%	.87	.40	.28	.06
10%	.91	.67	.46	.13	10%	.91	.54	.41	.13
Model B: $r = (1, 1; 1, 1)$									
$\rho = 0$					$\rho = .5$				
		b=2	b=8				b=2	b=8	
5%	.81	.13	.84	5%	.82	.12	.84		
10%	.86	.21	.87	10%	.86	.21	.86		

The statistic g^* is also insensitive to correlation changes; the values of the power being .05 and .10 for 5 and 10 percent significance levels, respectively. As compared with the statistic R , g^* is slightly more powerful under model A but slightly less powerful under model B. However, g^* is clearly less powerful than g and is also more difficult to compute.

Acknowledgement: Thanks are due to the NSERC of Canada and McMaster University Research Board for research grants.

REFERENCES

Barnett, V. (1978). The study of outliers: purpose and model. *Applied Statistics*, 27, 242-250.

Barnett, V. (1979). Some outlier tests for multivariate samples. *South African Statistical Journal*, 13, 29-52.

Barnett, V. and Lewis, T. (1978). *Outliers in Statistical Data*. Wiley, New York.

- Cox, D. R. (1968). Notes on some aspects of regression analysis. *Journal of the Royal Statistical Society, Series A*, 131, 265-279.
- David, H. A. and Paulson, A. S. (1965). The performance of several tests for outliers. *Biometrika*, 52, 429-436.
- Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28, 81-124.
- Hawkins, D. M. (1977). Comment on "A new statistic for testing suspected outliers". *Communications in Statistics*, A6, 435-438.
- Healy, M. J. R. (1968). Multivariate normal plotting. *Applied Statistics*, 17, 157-161.
- Johnson, N. L., Nixon, E., Amos, D. E. and Pearson, E. S. (1963). Table of percentage points of Pearson curves. *Biometrika*, 50, 459-498.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Siotani, M. (1959). The extreme value of the generalized distances of the individual points in the multivariate normal sample. *Annals of the Institute of Statistical Mathematics*, 10, 183-208.
- Tietjen, G. L. and Moore, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers. *Technometrics*, 14, 583-597.
- Tiku, M. L. (1967). Estimating the mean and standard deviation from censored normal samples. *Biometrika*, 54, 155-165.
- Tiku, M. L. (1970). Monte Carlo study of some simple estimators in censored normal samples. *Biometrika*, 57, 207-210.
- Tiku, M. L. (1975). A new statistic for testing suspected outliers. *Communications in Statistics*, 4, 737-752.
- Tiku, M. L. (1977). Rejoinder: "Comment on 'A new statistic for testing suspected outliers'". *Communications in Statistics*, A6, 1417-1422.
- Tiku, M. L. (1978). Linear regression model with censored observations. *Communications in Statistics*, A7, 1219-1232.
- Tiku, M. L. (1980). Robustness of MML estimators based on censored samples and robust test statistics. *Journal of Statistical Planning and Inference*, 4, (123-143).
- Tiku, M. L. and Singh, M. (1980). Robust estimation of the variance-covariance matrix and its use in testing an assumed multivariate distribution. (submitted to *JSPI* for publication).
- Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhya, Series A*, 25, 407-426.

APPENDIX

Let X_a, X_{a+1}, \dots, X_b ($a = r_1 + 1, b = n - r_2$) be the Type II censored sample, obtained by arranging a random sample

x_1, x_2, \dots, x_n of size n in ascending order of magnitude and censoring the r_1 smallest and the r_2 largest observations.

The expressions for the robust (modified maximum likelihood; Tiku, 1967, 1980) estimators of the scale parameters of normal, exponential and uniform populations are as follows: *Normal*.

The MML (modified maximum likelihood) estimator of σ is given by

$$\hat{\sigma} = \{B + \sqrt{B^2 + 4AC}\} / 2\sqrt{A(A-1)}, \quad (5)$$

where $A = n - r_1 - r_2$, $B = r_2 \alpha_2 X_b - r_1 \alpha_1 X_a - (r_2 \alpha_2 - r_1 \alpha_1)K$,

and $C = \sum_{i=a}^b X_i^2 + r_2 \beta_2 X_b^2 - r_1 \beta_1 X_a^2 - m K^2$;

$m = n - r_1 - r_2 + r_2 \beta_2 - r_1 \beta_1$, $K = (\sum_{i=a}^b X_i + r_2 \beta_2 X_b - r_1 \beta_1 X_a) / m$; if $q_1 = r_1/n = q$

then $\alpha_1 = \alpha$ and $\beta_1 = -\beta$, and if $q_2 = r_2/n = q$ then $\alpha_2 = \alpha$ and $\beta_2 = \beta$. The values of α and β are given by Tiku (1967, Table 1). For $n \geq 10$, however, the values of α and β are obtained from the following equations (Tiku, 1970):

$$\beta = -f(t)\{t - f(t)/q\}/q \quad \text{and} \quad \alpha = \{f(t)/q\} - \beta t,$$

where t is determined by the equation $\Phi(t) = 1 - q$; Φ is the standard normal cdf.

The ordinary estimator of σ is given by $s = \sqrt{\{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)\}}$, that is, the estimator $\hat{\sigma}$ with $r_1 = r_2 = 0$. The distribution of $\{(n - r_1 - r_2 - 1)/(n-1)\}(\hat{\sigma}/s)$ is approximately Beta $B(n - r_1 - r_2 - 1, r_1 + r_2)$; see Tiku (1975, p. 747).

Exponential. The MML estimator of σ is given by

$$\hat{\sigma} = \left\{ \sum_{i=a}^b X_i + r_2 X_b - (n - r_1) X_a \right\} / (n - r_1 - r_2 - 1); \quad (6)$$

the ordinary estimator is given by $s = (\sum_{i=1}^n X_i - n X_1) / (n-1)$. The distribution of $\{(n - r_1 - r_2 - 1)/(n-1)\}(\hat{\sigma}/s)$ is exactly Beta $B(n - r_1 - r_2 - 1, r_1 + r_2)$; see Tiku (1975, p. 748).

Uniform. The MML estimator of σ is given by

$$\hat{\sigma} = (X_b - X_a) / (n - r_1 - r_2 - 1); \quad (7)$$

the ordinary estimator is given by $s = (X_n - X_1)/(n-1)$. The distribution of $\{(n-r_1-r_2-1)/(n-1)\}(\hat{\sigma}/s)$ is exactly Beta $B(n-r_1-r_2-1, r_1+r_2)$; see Tiku (1975, p. 749).

If $\rho_{k\ell} = 0$ for all k and ℓ ($k \neq \ell$), then the p ratios $\hat{\sigma}_k/s_k$, $k = 1, 2, \dots, p$, in the expression (3) are mutually independent for the multivariate normal and certain types of multivariate exponential and uniform populations, for example Gumbel's bivariate exponential family (Johnson and Kotz, 1972, p. 261).

$$e^{-x-y-\theta xy} \{(1+\theta x)(1+\theta y)-\theta\}; \quad x, y > 0. \quad (8)$$

For the population (8) with $\theta = 0$ (i.e. correlation $\rho = 0$), the distribution of $hg = \{(n-r_1-1)(n-r_2-1)/(n-1)^2\}(\hat{\sigma}_1/s_1)(\hat{\sigma}_2/s_2)$ is exactly the same as that of the product of two independent Beta variates $B(n-r_1-1, r_1)$ and $B(n-r_2-1, r_2)$; $r_1 = r_{11} + r_{12}$ and $r_2 = r_{21} + r_{22}$. Let then u be the product of m independent Beta variates $B(n-r_i-1, r_i)$, $i = 1, 2, \dots, m$; $m \leq p < n$. The moments of u are easy to work out and the first four moments can be used to fit Pearson curves from Johnson *et al.* (1963) tables; these curves give accurate approximations to the percentage points of numerous distributions (Johnson *et al.*). Expressions for the probability density function $f(u)$ of u and its probability distribution function $F(u)$ can also be worked out. For example, we have the following expressions:

$$\underline{m = 2 \text{ and } r_1 = r_2 = 1}$$

$$f(u) = -(n-2)^2 (\log u) u^{n-3}, \quad 0 < u < 1,$$

$$F(u) = u^{n-2} \{1 - (n-2) \log u\},$$

$$\underline{m = 2 \text{ and } r_1 = r_2 = 2}$$

$$f(u) = (n-2)^2 (n-3)^2 \{2(u-1) - (1+u) \log u\} u^{n-4}, \quad 0 < u < 1,$$

$$F(u) = u^{n-3} [-(n-3)(n-2)\{(n-2) + (n-3)u\} (\log u) \\ + (n-2)^2 (7-2n) + (n-3)^2 (2n-3)u].$$

$$\underline{m = 3 \text{ and } r_1 = r_2 = 1}$$

$$f(u) = [(n-2)^3 (\log u)^2 u^{n-3}] / 2, \quad 0 < u < 1,$$

$$F(u) = u^{n-2} [1 + \{1 - (n-2)(\log u)\}^2]/2.$$

$$\underline{m = 4 \text{ and } r_1=r_2=1}$$

$$f(u) = -[(n-2)^4 (\log u)^3 u^{n-3}]/6,$$

$$F(u) = -u^{n-2}\{(n-2)^3(\log u)^3 - 3(n-2)^2(\log u)^2 + 6(n-2)\log u - 6\}/6.$$

The above equations were used to obtain the exact percentage points d given in Table 3; Table 3 also gives the simulated (based on 100,000/n Monte Carlo runs) values of the probability $P(g \leq d)$ for the bivariate exponential and normal populations. It is clear that the above equations provide reasonable approximations. Incidentally, the Pearson curves (Johnson *et al.*, 1963) based on the first four moments of u produce (even with linear interpolation in Johnson *et al.* tables) exactly the same values of d as those given in Table 3.

TABLE 3: Simulated values of the probability $P(g \leq d)$, d being the exact* 5 and 10 percent points.

		Bivariate exponential**			Bivariate normal				
		ρ			ρ				
n	d	0.0	0.5	0.8	-.8	-.5	0	.5	.8
$r = (0, 1; 0, 1)$									
10	.699	.052	.057	.075	.038	.041	.044	.048	.058
	.778	.104	.105	.119	.084	.090	.094	.095	.108
20	.856	.050	.055	.069	.035	.041	.042	.046	.058
	.898	.104	.101	.113	.083	.084	.090	.094	.102
40	.929	.051	.056	.068	.034	.039	.037	.037	.042
	.951	.104	.102	.116	.084	.086	.081	.080	.084
$r = (0, 1; 1, 0)$									
10	.699	.050	.047	.047	.055	.045	.043	.046	.043
	.778	.099	.097	.092	.099	.086	.092	.090	.084
20	.856	.048	.051	.046	.052	.042	.042	.039	.036
	.898	.098	.101	.096	.093	.087	.088	.083	.080
40	.929	.050	.046	.044	.048	.045	.042	.036	.034
	.951	.096	.098	.104	.094	.088	.091	.079	.080
$r = (1, 1; 1, 1)$									
10	.584	.050	.054	.067	.063	.051	.044	.052	.065
	.675	.100	.101	.117	.115	.102	.101	.103	.116
20	.802	.050	.056	.067	.055	.044	.039	.044	.058
	.852	.104	.103	.108	.104	.088	.088	.094	.107
40	.903	.052	.055	.064	.043	.034	.037	.036	.044
	.930	.102	.102	.122	.096	.086	.086	.085	.092

* Exact percentage points of the distribution of u/h , u being the product of two independent Beta variates $B(n-r_1-1, r_1)$ and $B(n-r_2-1, r_2)$, and $h = (n-r_1-1)(n-r_2-1)/(n-1)^2$; $r_1 = r_{11} + r_{12}$ and $r_2 = r_{21} + r_{22}$.

** This is Moran's bivariate exponential family (Johnson and Kotz, 1972, p. 267) which is easy to generate; for this family correlation $\rho \geq 0$.

[Received April 1980. Revised August 1980]

A NORMAL APPROXIMATION FOR THE MULTIVARIATE LIKELIHOOD RATIO STATISTICS

GOVIND S. MUDHOLKAR*

Department of Statistics
University of Rochester
Rochester, New York 14627 USA

MADHUSUDAN C. TRIVEDI

Pennwalt Corporation
Pharamaceutical Division
P. O. Box 1710
Rochester, New York 14603 USA

SUMMARY. For many multivariate hypotheses, under the normality assumptions, the likelihood ratio tests are optimal in the sense of having maximal exact slopes. The exact distributions needed for implementing these tests are complex and their tabulation is limited in scope and accessibility. In this paper, a method of constructing normal approximations to these distributions is described, and illustrated using the problems of testing sphericity and independence between two sets of variates. The normal approximations are compared with well-known competing approximations and are seen to fare well.

KEY WORDS. Sphericity, independence between two sets of variates.

*Research supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF under Grant No. AFOSR-77-3360. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

1. INTRODUCTION

For most testing of hypothesis problems in multivariate analysis, under the normality assumption, several reasonable solutions of comparable merit exist. These include the tests resulting from the union-intersection principle, the class of likelihood ratio criteria and *ad hoc* statistics such as Bartlett-Pillai trace for MANOVA. The Neyman-Pearson theory provides some information on the operating characteristics of these procedures, but does not indicate any of the contenders as superior. However, as demonstrated by Hsieh (1979), the likelihood ratio tests for many of the multivariate hypotheses have maximal exact slopes, i.e., they are asymptotically optimal according to Bahadur's (1967) method of comparing tests. From a practical standpoint the null distributions of the likelihood ratio statistics or of their competitors, are of crucial importance. These distributions, where available, are complex, their tables are generally limited in scope and not often accessible. Moreover, the tabulations concern only selected percentiles and are inadequate for computing the p-values needed in practice. The pragmatic approach to such distribution problems from early days (e.g., Neyman and Pearson, 1931) is to seek reasonably accurate and convenient approximations to the distributions.

The principal methods of approximating a likelihood ratio use the fact that, in large samples, its distribution is approximately of Pearson type I form and that of its negative logarithm is of type III, i.e., chi-square, form. Nayer (1936) following a suggestion by Neyman and Pearson (1931) used the moments to approximate the percentiles for testing the homogeneity of variances in this manner. Bishop (1939), on the other hand, obtained empirical expressions for the parameters for a type I approximation by passing the intermediate stage of computing the moments. Bartlett (1937) pursuing the asymptotic chi-square character of a negative multiple of log-likelihood ratio, pointed out by Neyman and Pearson (1931), used moments to approximate it by a scaled chi-square variable for samples of moderate size. This approximation deteriorates as the size of the problem, as measured by the dimension of the multivariate normal distribution or by the number of populations in the problem increases, or when the effective sample size is small. A comprehensive investigation of various approximations was conducted by Box (1949), in which he introduced new widely known and used asymptotic chi-square series approximations for the distributions of likelihood ratios. Box studied his series approximations, in the context of two multivariate problems, comparing them with the exact distributions and with several other approximations including one based on the F-distribution.

The purpose of this essay is to describe a method for constructing a Gaussian approximation to the null distribution of the likelihood ratio, and to demonstrate its efficacy and relevance in testing multivariate hypotheses. The normal approximation is outlined in Section 2. It is illustrated using two common multivariate problems, namely testing independence of two sets of variates and testing the sphericity hypotheses. Section 3 contains the likelihood ratio statistics for the two problems together with current approximations for their null distributions. These approximations are then numerically compared with the new normal approximation in Section 4.

2. A NORMAL APPROXIMATION FOR THE LIKELIHOOD RATIO Λ

Let Y_1, Y_2, \dots, Y_n be a sequence of asymptotically normally distributed nonnegative random variables. The convergence of the distribution of Y_n to normality can be accelerated by approximately symmetrizing it with a transformation as follows:

Let $\kappa_r = \kappa_r(n)$, $r = 1, 2, \dots$, denote the cumulants of $Y = Y_n$ and suppose that $\kappa_1 \rightarrow \infty$ and $\kappa_r/\kappa_1 = \phi_r$, $r \geq 2$, are bounded as $n \rightarrow \infty$. Then using the Taylor series it is easy to obtain the following asymptotic expansion for the expectation $E(Y/\kappa_1)^h$ of a power of Y as

$$\mu'_1(h) = 1 + \frac{h(h-1)\phi_2}{2\kappa_1} + \frac{h(h-1)(h-2)}{24\kappa_1^2} [4\phi_3 + 3(h-3)\phi_2^2] + O(\kappa_1^{-3}). \quad (1)$$

For this the r th moment of $(Y/\kappa_1)^h$ can be obtained by substituting (rh) for h in (1). The following central moments of $(Y/\kappa_1)^h$ are then obtained in a routine manner:

$$\mu_2(h) = \frac{h^2\phi_2}{\kappa_1} + \frac{h^2(h-1)}{2\kappa_1^2} [2\phi_3 + (3h-5)\phi_2^2] + O(\kappa_1^{-3}), \quad (2)$$

$$\mu_3(h) = \frac{h^3}{\kappa_1} [\phi_3 + 3(h-1)\phi_2^2] + O(\kappa_1^{-3}), \quad (3)$$

$$\mu_4(h) = \frac{3h^4\phi_2^2}{\kappa_1^2} + O(\kappa_1^{-3}).$$

Since Y is asymptotically normally distributed as $n \rightarrow \infty$, by Mann-Wald (1943) theorem so is an appropriately normalized $(Y/\kappa_1)^h$. This convergence to normality is accelerated if h is chosen so that the leading term in the expansion (3) for $\mu_3(h)$ vanishes. This value h_0 of h which approximately symmetrizes $(Y/\kappa_1)^h$ is obtained from (3) as

$$h_0 = 1 - \kappa_1 \kappa_3 / (3\kappa_2^2).$$

The distribution of $(Y/\kappa_1)^{h_0}$ may be approximated by the normal distribution with mean $\mu_1'(h_0)$ and variance $\mu_2(h_0)$ given in (1) and (2), respectively. That is,

$$\Pr(Y \leq y) \approx \Phi[\{(y/\kappa_1)^{h_0} - \mu_1'(h_0)\} / \sigma(h_0)], \quad (4)$$

where $\sigma^2(h_0) = \mu_2(h_0)$ is given by (2).

It is well known, e.g., see Anderson (1958) or Srivastava and Khatri (1979), that for many likelihood ratio statistics Λ appearing in multivariate analysis under the normality assumption, $U = \Lambda^{2/N}$ is distributed as a product $\prod X_i$ of independent beta variates X_i , $i = 1, 2, \dots, k$, distributed according to

$B(X_i; a_i, b_i)$, where N is the number of observations. Equivalently, we have $-\log U = \sum_{i=1}^k (-\log X_i)$ in distribution.

Now, it can be shown that, as a_i and $b_i \rightarrow \infty$, $-\log X_i$ converges in law to normality. Hence, it is possible to construct a normal approximation for U as described above. Towards this end we need the cumulants of $-\log X_i$. The moment generating function of $-\log X_i$ is easily seen to be $M(t) = B(a_i - t, b_i) / B(a_i, b_i)$. Hence, the cumulant generating function is

$$K(t) = \log[\Gamma(a_i + b_i) / \Gamma(a_i)] - \log[\Gamma(a_i + b_i - t) / \Gamma(a_i - t)].$$

Differentiating and using $\Psi(Z) = \frac{d}{dZ} \log \Gamma(Z)$, the r th cumulant of $-\log X_i$ is

$$C_{ri} = (-1)^r [\Psi^{(r-1)}(a_i) - \Psi^{(r-1)}(a_i + b_i)].$$

But $\Psi'(Z) = - \sum_{j=0}^{\infty} (Z+j)^{-1}$, giving

$$C_{ri} = (r-1)! \left[\sum_{j=0}^{m-1} (a_i+j)^{-r} + \sum_{j=m}^{\infty} \{(a_i+j)^{-r} - (a_i+j+v)^{-r}\} \right], \quad (5)$$

where m denotes the largest integer in b_i and $v = b_i - m$. The cumulants of $-\log U'$ obtained using (5) are,

$$\begin{aligned} \kappa_r(U') = (r-1)! & \left[\sum_{i=1}^k \sum_{j=0}^{m-1} (a_i+j)^{-r} \right. \\ & \left. + \sum_{i=1}^k \sum_{j=m}^{\infty} \{(a_i+j)^{-r} - (a_i+j+v)^{-r}\} \right]. \end{aligned} \quad (6)$$

If b_i is an integer then the second sum in (6) vanishes and

$$\kappa_r(U') = (r-1)! \sum_{i=1}^k \sum_{j=0}^{m-1} (a_i+j)^{-r}. \quad (7)$$

From (7) we observe that as either k or b_i or both $\rightarrow \infty$, κ_1 diverges, but κ_r , $r > 1$, are bounded. That is, $\phi_r = \kappa_r/\kappa_1 \rightarrow 0$. Hence, it is possible to construct the normal approximation to the distribution of Λ as described above. Thus, from (4) we get

$$\Pr(\Lambda \geq \lambda) \approx \phi\{[(\lambda'/\kappa_1)^{h_0} - \mu_1'(h_0)]/\sigma(h_0)\}, \quad (8)$$

where $\lambda' = -2(\log \lambda)/N$. The $100(1-\alpha)\text{th}$ percentile $\Lambda_{1-\alpha}$ can be approximated as

$$\Lambda_{1-\alpha} \approx \kappa_1 [Z_{\alpha} \sigma(h_0) + \mu_1'(h_0)]^{1/h_0}, \quad (9)$$

where Z_{α} denotes the $100\alpha\text{th}$ percentile of the standard normal variate.

3. TWO APPLICATIONS IN MULTIVARIATE ANALYSIS

The normal approximation derived in the previous section is now illustrated and later examined in the context of the multivariate problems of testing independence between two sets of normal variates and testing the sphericity hypothesis.

3.1 Independence Between Two Sets. Let $\tilde{X}' = (X_1, X_2, \dots, X_{p_1})$ and $\tilde{Y}' = (Y_1, Y_2, \dots, Y_{p_2})$, $p_1 \leq p_2$, $p_1 + p_2 = p$, be jointly

normally distributed with $\text{Var}(\tilde{X}) = \Sigma_{11}$, $\text{Var}(\tilde{Y}) = \Sigma_{22}$ and $\text{Cov}(\tilde{X}, \tilde{Y}) = \Sigma_{12}$. The hypothesis of independence between \tilde{X} and \tilde{Y} is $H_0: \Sigma_{12} = 0$. If S is the usual estimate of Σ based on a sample of size N , then the likelihood ratio statistic for testing H_0 is

$$\Lambda = [|S| / (|S_{11}| |S_{22}|)]^{N/2},$$

where S_{11} and S_{22} are the submatrices of S corresponding to Σ_{11} and Σ_{22} , respectively. The exact distribution of the statistic Λ is given, and tabulated for some values of p_1 and p_2 , by several authors (e.g., see Krishnaiah, 1979, and Consul, 1967a). Among various approximations proposed for the null distribution of Λ two are well known and widely used in statistical packages such as BMDP (see Engelman *et al.*, 1977). These are (i) the chi-square series approximation due to Box (1949) and (ii) the F approximation due to Rao (1948).

Box-Approximation. Let $w = p_1 p_2$, $m = N - (p_1 + p_2 + 3)/2$, $\gamma_2 = w(p_1^2 + p_2^2 - 5)/48$, $\gamma_4 = \gamma_2^2/2 + w\{3(p_1^4 + p_2^4) + 10w^2 - 50(p_1^2 + p_2^2) + 159\}/1920$. Then,

$$\begin{aligned} \Pr(-m \log U \leq z) &\approx \Pr(\chi_w^2 \leq z) + \gamma_2 \{\Pr(\chi_{w+4}^2 \leq z) - \Pr(\chi_w^2 \leq z)\}/m^2 \\ &\quad + [\gamma_4 \{\Pr(\chi_{w+8}^2 \leq z) - \Pr(\chi_w^2 \leq z)\} \\ &\quad - \gamma_2^2 \{\Pr(\chi_{w+4}^2 \leq z) - \Pr(\chi_w^2 \leq z)\}]/m^4 + O(N^{-6}), \end{aligned} \quad (10)$$

where χ_k^2 denotes a chi-square variable with k degrees of freedom and $U = \Lambda^{2/N}$.

Rao-Approximation. Let $m' = N - (p_1 + p_2 + 3)/2$, $L = (p_1 p_2 - 2)/4$, $s = \sqrt{[(p_1^2 p_2^2 - 4)/(p_1^2 + p_2^2 - 5)]}$. Then,

$$Q = (m's - 2L)(1 - U^{1/s})/(p_1 p_2 U^{1/s}), \quad (11)$$

has an F -distribution with $p_1 p_2$ and $m's - 2L$ degrees of freedom.

Now, it is well known that (e.g., see Anderson, 1958, p. 236) under H_0 the likelihood ratio statistic Λ satisfies the equivalence $\Lambda^{2/N} = U = \prod X_i$ in law, where X_i ($i = 1, 2, \dots, p_2$) are independently distributed according to beta distributions $B\{X_i; (N-p_1-i)/2, p_1/2\}$. The normal approximation developed in the previous section can be specialized in this case by taking $k = p_2$, $a_i = (N-p_1-i)/2$, $b_i = p_1/2$ in the expressions (6) for the cumulants, (8) for the probabilities, and (9) for the percentiles of Λ .

3.2 Testing the Sphericity Hypothesis. Let $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_N$ be a random sample from a p -variate normal population with mean $\tilde{\mu}$ and covariance matrix $\tilde{\Sigma}$. The hypothesis that the p components of the random vector \tilde{X} are independent with the same variance, i.e., $H_0: \tilde{\Sigma} = \sigma^2 \tilde{I}_p$, $\sigma^2 > 0$ unknown, is known as the sphericity hypothesis. The hypothesis also arises in the analysis of data from experiments consisting of repeated measurements. In these experiments, the measurements on a subject are assumed to have compound symmetry, i.e., have the same variances and same correlations. The problem of testing the hypothesis of compound symmetry $H_0: \tilde{\Sigma} = \sigma^2(\rho \tilde{I} + (1-\rho)\tilde{I})$ for the covariance structure of $(p+1)$ repeated measurements \tilde{Y} can be reduced to the sphericity hypothesis by an orthogonal transformation $\tilde{Y}' \rightarrow \tilde{Y}'(1/\sqrt{p+1}):T_1$ where $\tilde{1}$ is the vector of 1's. \tilde{Y} satisfies compound symmetry if and only if $\tilde{X} = T_1 \tilde{Y}$ satisfies the sphericity hypothesis. The likelihood ratio criterion for the sphericity hypothesis was proposed by Mauchly (1940) as

$$U = \Lambda^{2/N} = |S| [(tr S)/p]^{-p},$$

where S is the covariance matrix of the sample of size N . He also derived its null distribution for $p = 2$. The exact null distribution of U for $p = 3, 4$, and 6 was obtained by Consul (1967b). The 5% and 1% points for $p = 4(1)10$ were given by Nagarsanker and Pillai (1973). The series approximation due to Box can be expressed in this case as follows.

Box-Approximation. Let $e = p(p+1)/2 - 1$, $f = n - (2p^2 + p + 2)/(6p)$ and $g = (p+2)(p-1)(p-2)(2p^3 + 6p^2 + 3p + 2)/(288p^2)$ for $n = N-1$. Then,

$$\begin{aligned} \Pr(-f \log U \leq z) &\approx \Pr(\chi_e^2 \leq z) + g\{\Pr(\chi_{e+4}^2 \leq z) \\ &\quad - \Pr(\chi_e^2 \leq z)\}/f^2 + o(f^{-3}). \end{aligned} \quad (12)$$

It is well known (e.g. see Srivastava and Khatri, 1979) that the distribution of U under H_0 is the same as that of the product $\prod X_i$, where X_i ($i = 1, 2, \dots, p-1$) are independent beta random variables distributed according to $B\{x_i; (n-i)/2, i(p+2)/(2p)\}$, $n = N-1$. Again, we can obtain the cumulants of $U' = -2(\log U)/N$ using (6) with $a_i = (n-i)/2$, $b_i = i(p+2)/(2p)$ and $k = p-1$. Hence, the probabilities and the percentiles of the likelihood ratio Λ may be obtained from (8) and (9), respectively.

4. NUMERICAL COMPARISONS

The quality of the normal approximations for the two multivariate likelihood ratio statistics discussed in the previous section and the other two approximations, was examined by computing the probabilities corresponding to the tabulated percentiles of the statistics. Thus, in the case of the null distribution of Λ for testing independence, the approximation due to Box (10), due to Rao (11), and the normal approximation given in Section (3.1) were used to compute the probabilities corresponding to all 5% and 1% points of Λ given in Pearson and Hartley (1972, p. 99 and 333). Similarly, in case of the sphericity problem, all percentiles given by Nagarsanker and Pillai (1973) were used to examine the approximation due to Box given by (12) and the relevant normal approximation. In both cases, the series approximation due to Box was used in two steps: 1) only the first term; and 2) all terms given in (10) and (12). Also the percentiles approximated using the normal approximations were compared with the competing approximations using the first term of the Box series and the F-approximation. A selection of errors, i.e., $(\text{Approximation} - \text{Exact value}) \times 10^5$, in various cases is presented in Tables 1 and 2.

Conclusions. Let New, Rao, Box 1 and Box 3 denote the normal approximation, the F-approximation due to Rao, the first term approximation due to Box and the three term approximation due to Box, respectively. From Tables 1 and 2 it may be observed that: (i) Rao, Box 1 and Box 3 have errors in the second through fifth decimal place, they are especially large for small N and decreasing rapidly as N increases. The normal approximation has errors in the fourth or fifth decimal place. (ii) As p_1, p_2

TABLE 1: Errors of approximations for the likelihood ratio statistic for testing sphericity

P	N	λ	$\alpha = .05$				λ	$\alpha = .01$					
			Errors*					Errors*					
			Probability					Probability					
			Percentile	Box1	New	Box2		Percentile	Box1	New	Box2		
4	10	0.09739	16	467	-21	-539	-57	0.05010	-22	382	10	-173	-25
	15	0.25350	35	236	-29	-190	-10	0.17210	-41	267	9	-62	-4
	20	0.37720	43	120	-33	-97	-5	0.28670	-44	166	8	-31	-1
	30	0.53900	46	37	-38	-40	-2	0.45310	-38	76	7	-12	0
5	10	0.03110	1	444	-7	-1171	-236	0.01361	-5	281	7	-350	-96
	15	0.13780	7	334	-10	-393	-29	0.08685	-19	287	7	-124	-14
	20	0.24820	17	201	-17	-194	-7	0.17970	-23	188	6	-63	-5
	30	0.41640	28	79	-27	-78	-2	0.34020	-29	58	6	-23	0
7	10	0.00094	0	104	20	-3352	-1789	0.00025	0	43	2	-822	-566
	15	0.02712	0	297	-4	-1158	-228	0.01444	-3	207	4	-330	-86
	20	0.08446	5	282	-12	-562	-63	0.05514	-5	234	3	-165	-24
	30	0.21780	15	163	-21	-216	-13	0.16770	-4	153	1	-65	-5

* Error in Probability = (Approx. value - α) $\times 10^5$.Error in Percentile = (Approx. value - λ) $\times 10^5$.

TABLE 2. *Errors of the approximations for the likelihood ratio statistic for testing independence between two sets of variates*

$\alpha = .05$										
P1	P2	N	λ	Errors*						
				Percentiles			Probabilities			
				New	Rao	Box1	New	Rao	Box1	Box3
3	8	12	0.00001	0	0	67	-19	-1537	-4991	-4744
	8	19	0.04107	4	7	698	-17	-24	-1776	-50
	22	30	0.00107	0	1	356	0	-143	-4936	-3893
	22	37	0.01620	0	0	760	-4	0	-3842	-725
5	8	18	0.00217	0	5	161	2	-218	-3340	-485
	8	25	0.03411	1	3	352	-8	-17	-1355	-16
	16	26	0.00019	0	0	50	12	-420	-4759	-2732
	16	33	0.00568	0	1	213	-6	-48	-3160	-316
7	8	19	0.00021	0	1	37	-9	-618	-4225	-1437
	8	23	0.00356	0	3	133	-3	-122	-2657	-199
	10	21	0.00007	0	0	18	25	-746	-4531	-2059
	10	25	0.00157	0	2	81	9	-155	-3153	-358
$\alpha = .01$										
3	8	12	0.00000	0	0	17	6	-478	-999	-997
	8	19	0.02261	0	7	514	0	-9	-487	-29
	22	30	0.00043	0	0	209	0	-44	-997	-935
	22	37	0.00990	-2	-1	572	6	5	-872	-303
5	8	18	0.00085	0	3	85	3	-72	-803	-223
	8	25	0.02086	0	6	270	-2	-11	-374	-15
	16	26	0.00007	0	0	25	-10	-143	-985	-784
	16	33	0.00327	0	0	147	3	-11	-753	-147
7	8	19	0.00007	0	0	16	1	-192	-932	-513
	8	23	0.00174	0	2	81	-1	-42	-665	-102
	10	21	0.00002	0	0	8	3	-234	-965	-659
	10	25	0.00075	0	1	48	5	-48	-756	-164

* Error in Probability = (Approx. value - α) $\times 10^5$.

Error in Percentile = (Approx. value - λ) $\times 10^5$.

or p increases, errors due to Rao, Box 1 and Box 3 increase while those due to the normal approximation either decrease or maintain the same level. Overall, the normal approximation is superior for small N and is comparable with the others when N is large.

REFERENCES

- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- Bahadur, R. R. (1967). An optimal property of the likelihood ratio statistics. *Proceedings of Fifth Berkeley Symposium Mathematical Statistics and Probability*. University of California Press. Pages 13-26.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of Royal Statistical Society, Series A*, 160, 268.
- Bishop, D. J. (1939). On the comprehensive test of the homogeneity of variances and covariances of multivariate problems. *Biometrika*, 31, 31-55.
- Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.
- Consul, P. C. (1967a). On the exact distribution of likelihood ratio criteria for testing independence of sets of variates under the null hypothesis. *Annals of Mathematical Statistics*, 38, 1160-1169.
- Consul, P. C. (1967b). On the exact distribution of the criterion W for testing sphericity in a p -variate normal distribution. *Annals of Mathematical Statistics*, 38, 1170-1174.
- Engelman, L., Frane, J. W., and Jennrich, R. I. (1977). *Biomedical Computer Programs P-Series*. University of California Press, Berkeley, California.
- Hsieh, H. K. (1979). On asymptotic optimality of likelihood ratio tests for multivariate normal distributions. *Annals of Statistics*, 7, 592-598.
- Krishnaiah, P. R. (1979). *Some Recent Developments on The Real Multivariate Distributions*. North Holland Co., New York.
- Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *Annals of Mathematical Statistics*, 14, 217-226.
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n -variate distribution. *Annals of Mathematical Statistics*, 11, 204-209.
- Nagarsanker, B. N. and Pillai, K. C. S. (1973). Distribution of the sphericity test criterion. *Journal of Multivariate Analysis*, 3, 226-235.
- Nayer, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage

- limits. *Statistical Research Memorandum*, 1, 38. University of London.
- Neyman, J. and Pearson, E. S. (1931). On the problem of k samples. *Bulletin International Academy Cracovie*, A, 6, 460.
- Pearson, E. S. and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, 2. Cambridge University Press, New York.
- Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika*, 35, 58-79.
- Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North Holland Co., New York.

[Received June 1980. Revised October 1980]

EXPLICIT ACCURATE APPROXIMATIONS FOR FITTING THE PARAMETERS OF L_U

K. O. BOWMAN

Computer Sciences Division
Union Carbide Corporation
Nuclear Division
Oak Ridge, Tennessee 37830 USA

L. R. SHENTON

Office of Computing Activities
University of Georgia
Athens, Georgia 30602 USA

SUMMARY. The Tadikamalla-Johnson unlimited system based on the Logistic density may be fitted through the equivalence of the skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) in the model and distribution approximated. One-shot accurate approximations compact enough for small calculators and covering an extensive domain are described.

KEYWORDS. approximations, Johnson's system, least squares, moments, transformations, rational fraction.

1. GENESIS OF APPROXIMATIONS

L_U has been described recently by Tadikamalla and Johnson (1979) and consists of the transformation

$$Z = \gamma + \delta \sinh^{-1} Y, \quad (1)$$

where Z has the logistic density

$$f_Z(z) = \exp(z) \{1 + \exp(z)\}^{-2} \quad (-\infty < z < \infty)$$

and simple cumulative function

$$F_Z(z) = \{1 + \exp(-z)\}^{-1}.$$

They give the first four non-central moments as

$$\mu_1' = -\theta \operatorname{cosec} \theta \sinh \Omega, \quad (\theta = \pi/\delta)$$

$$\mu_2' = \theta \operatorname{cosec} 2\theta \cosh 2\Omega - 1/2,$$

$$\mu_3' = -(3/4) \theta (\operatorname{cosec} 3\theta \sinh 3\Omega - \operatorname{cosec} \theta \sinh \Omega),$$

$$\mu_4' = (1/2) \theta (\operatorname{cosec} 4\theta \cosh 4\Omega - 2 \operatorname{cosec} 2\theta \cosh 2\Omega) + 3/8,$$

from which the central moments can be derived and the skewness ($\sqrt{\beta_1} = \mu_3/\mu_2^{3/2}$) and kurtosis ($\beta_2 = \mu_4/\mu_2^2$).

The parameters γ, δ in (1) are determined from

$$\sqrt{\beta_1}(\gamma, \delta) = \sqrt{\beta_1}, \quad \beta_2(\gamma, \delta) = \beta_2 \quad (2)$$

for a given $\sqrt{\beta_1}, \beta_2$, assuming a solution exists. The solution is completed for a working variate X by setting $Y = (X - \xi)/\lambda$, the parameters ξ, λ being determined from the equivalence of location and scale. Since (2) requires the algorithmic solution of bivariate equations beyond the capacity of calculators, with a consequent resort to tabulation (and possibly interpolation problems), other approaches seem worthwhile.

In deriving an algorithm for the corresponding problem with S_U , Johnson (1965) noted the linearity of δ -contours in the (β_1, β_2) plane and also the stability of the function $\{\beta_2 - \frac{1}{2}(\omega^4 + 2\omega^2 + 3)\}/\beta_1$. Bowman and Shenton (1979) have used the latter, approximated by a rational fraction in β_1 and β_2 , to determine $\omega (= \exp 1/\delta^2)$ with very acceptable accuracy over a much larger domain than the Johnson tabulations, the approximation being exact for $\beta_1 = 0$.

We follow this pattern for L_U , and are led to a one-shot solution for δ , and a solution for γ derived from this.

2. AN EXPLICIT APPROXIMATION FOR θ

We first of all considered the solution when $\beta_1 = 0$ so that $\Omega = \gamma/\delta = 0$, and $\theta = \pi/\delta$ is to be found (see Tadikamalla and Johnson, 1979). The equation is

$$\beta_2 = f(\theta), \quad (3)$$

where

$$f(\theta) = \left\{ \frac{1}{2} \theta (\operatorname{cosec} 4\theta - 2 \operatorname{cosec} 2\theta) + \frac{3}{8} \right\} / \left(\theta \operatorname{cosec} 2\theta - \frac{1}{2} \right)^2.$$

As is the case for S_U (Bowman and Shenton, 1979), we now consider the effect of a small skewness, and entertain an expansion of $\beta_2 - f(\theta)$ as $a_1(\theta)\beta_1 + a_2(\theta)\beta_1^2 + \dots$. The next step was to consider contours of $g(\theta) = [\beta_2 - f(\theta)]/\beta_1$; these turned out to be almost linear (Figure 1), with almost constant gradient with respect to β_2 , in a domain $\beta_1 = 0$, $4.2 < \beta_2 < 25$, and the log-logistic line L_L (this L_L line is almost linear provided $\sqrt{\beta_1}$ is less than 2 approximately). However, even if a simple approximation to $g(\theta)$ was found, we should still have a transcendental equation to solve for θ . A closer look at (3) suggests the possibility of an expansion for θ^2 (not θ) as a polynomial in β_1 and $\beta_2^* = \beta_2 - 4.2$; for $\beta_1 = 0$, and $\beta_2^* = 0$ is an extreme point on L_L for which $\theta = 0$. It is plausible then to consider the expansion

$$\theta^2 = a_{10}\beta_1 + a_{01}\beta_2^* + a_{20}\beta_1^2 + \dots \quad (4)$$

However we point out that L_U is such that the 4th moment involves $\operatorname{cosec} 4\theta$, so that since $\theta < \pi/4$ for the existence of β_2 , this shows that the extreme value of $|\sqrt{\beta_1}|$ is

$$k = 2(12 - 6\pi + \pi^2)/[\sqrt{\pi}(4 - \pi)^{3/2}] \quad (5)$$

$$= 4.2847 \dots,$$

when $\beta_2 \rightarrow \infty$. Hence an improvement on (4) is

$$\theta^2 = \sum_{rs} a_{rs}^{(1)} \beta_1^r \beta_2^{*s} / \sum_{rs} a_{rs}^{(2)} \beta_1^r \beta_2^{*s}$$

with $a_{00}^{(1)} = 0$ and $a_{00}^{(2)} = 1$, which can remain finite as $\beta_2 \rightarrow \infty$, provided there are compensating terms in β_2^* in numerator and denominator. Finally then we use the approximation,

$$\theta_a^2 = G_1(\beta_1, \beta_2) = \pi_1(\beta_1, \beta_2^*)/\pi_2(\beta_1, \beta_2^*) \quad (6)$$

where, for $i=1,2$,

$$\pi_i(\beta_1, \beta_2^*) = \sum_{0 \leq r+s \leq 3} \sum_{rs} a_{rs}^{(i)} \beta_1^r \beta_2^{*s}$$

with $a_{00}^{(1)} = 0$ and $a_{00}^{(2)} = 1$.

The eighteen coefficients in (6) may be determined by using linearized least squares over a representative set of grid points for the triad $(\beta_1, \beta_2; \delta)$; we used 1000 points on a grid for which $\sqrt{\beta_1} = 0.01, 0.1, 0.2(0.2)3$; for β_2 we computed, for these β_1 values, the L_L value β_{2L} and used the integer part of $1 + \beta_{2L}$ incremented by unity up to $\beta_2 = 75$. The coefficients (Table 1) show that the cubic terms are not playing a dominant role unless β_1 and (or) β_2 are large.

As an obvious check on (6) one can assume β_1 small and $\beta_2 = 4.2$ approximately in the moment equations, and mathematically derive the approximation

$$\theta^2 \approx \{p(\beta_2 - 4.2) + q\beta_1\}/\{1 - r(\beta_2 - 4.2) + s\beta_1\}$$

where $p = 175/864$, $q = -425/1152$, $r = 35/648$, $s = 85/864$. In

TABLE 1: Rational fraction approximations for θ^2 .

rs	$\theta_a^2 = G_1(\beta_1, \beta_2)$		$\theta_a^{*2} = G_2(\beta_1, \beta_2)$	
	$a_{rs}^{(1)}$	$a_{rs}^{(2)}$	$a_{rs}^{(1)}$	$a_{rs}^{(2)}$
10	-3.689183814049335 (1)	-6.796170149361249 (1)	-3.683814287125019 (1)	-1.515253321453197 (1)
01	2.025419425875670 (1)	6.444506376151906 (1)	2.023584261173814 (1)	2.561517934295567 (1)
20	1.217331928564157 (1)	1.108627451888432 (1)	-7.164215421040600 (2)	-6.853092721414190 (2)
11	-1.828076112994866 (1)	-2.158471102836237 (1)	6.594403750391700 (2)	8.906614338276530 (2)
02	6.352884107647720 (2)	1.030076605210349 (1)	-1.478655283885235 (2)	-2.411331184683489 (2)
30	1.502210979590406 (5)	2.014132479989601 (6)		
21	1.154479372054198 (4)	8.474617911907420 (5)		
12	-1.081371626060427 (4)	-1.705393457376706 (4)		
03	3.731479112821334 (6)	6.049134822296229 (6)		

(Parenthetic entries taken negatively give the power of 10 by which the corresponding entry is to be multiplied.)

(6) $a_{10}^{(1)} = -0.368918$, $a_{01}^{(1)} = 0.202542$ compared to $q = -0.368924$, $p = 0.202546$, in excellent agreement.

3. THE SOLUTION FOR Ω

Knowing θ in (2) we solve for $t = \sinh \Omega$ from the kurtosis equations. Thus, defining $c_s = \operatorname{cosec} s\theta$ for simplicity and with $\theta = \theta_a$

$$\begin{aligned} a &= 2\theta c_2 - \theta^2 c_1^2, & b &= \theta c_2 - 1/2 \\ p &= 4\theta c_4 - 12\theta^2 c_1 c_3 + 12\theta^3 c_1^2 c_2 - 3\theta^4 c_1^4 \\ q &= 4\theta c_4 - 2\theta c_2 - 9\theta^2 c_1 c_3 + 6\theta^3 c_1^2 c_2 \\ r &= (1/2)\theta c_4 + 3/8 - \theta c_2, \end{aligned} \quad (7)$$

we have

$$t = [(2ab\beta_2 - q + \sqrt{H}) / (2p - 2a^2\beta_2)]^{1/2}$$

where $H = q^2 - 4pr + 4\beta_2 (ra^2 + pb^2 - qab),$

and $\Omega = \ln[t + \sqrt{(t^2 + 1)}].$

Since the kurtosis equation is now satisfied, the only source of error in the solution for Ω arises from θ_a .

Numerical results show that there is serious loss of accuracy in (7) when $\sqrt{\beta_1}$ is small. In this case it is advisable to use the equation for $\sqrt{\beta_1}$, namely

$$\sqrt{\beta_1} = t(At^2 + B)/(at^2 + b)^{3/2} \quad (8)$$

where $t = \sinh \Omega$, a and b are given in (7), and

$$A = 6\theta^2 c_1 c_2 - 2\theta^3 c_1^3 - 3\theta c_3$$

$$B = 3\theta^2 c_1 c_2 - 3\theta c_1/4 - 9\theta c_4/4 \quad (\theta < \theta_a).$$

The obvious method to solve (8) would be to express it as a cubic in t^2 and use an appropriate subroutine. However it turns out to be simpler to use the iterative scheme

$$t_n = \sqrt{\beta_1} (at_{n-1}^2 + b)^{3/2} / (At_{n-1}^2 + B), \quad n=1,2,\dots, \quad (9)$$

with $t_0 = \sqrt{\beta_1}$. This converges rapidly for small $\sqrt{\beta_1}$.

To complete the solution, having found values for θ and Ω , we have, defining the mean and variance of the working variate X as v_1' and v_2 respectively, with $\mu_2 = \mu_2' - \mu_1'^2$, $\lambda = \sqrt{(v_2/\mu_2)}$ and $\xi = v_1' - \lambda\mu_1'$.

4. DOMAIN OF VALIDITY

The solution (θ_a, Ω_a) in (6), (7), and (9) is designed for the domain $\beta_1 = 0$, $4.2 < \beta_2 \leq 75$ and a slightly modified L_L line. To define the latter, we use the approximation

$$\hat{\beta}_2 = (a_0 + a_1\sqrt{\beta_1} + a_2\beta_1 + a_3\beta_1^{3/2})/(k - \sqrt{\beta_1})$$

to define $\hat{\beta}_2$ on L_L . The coefficients are

$$a_0 = 17.64798988, \quad a_1 = -2.364654939,$$

$$a_2 = 7.598129943, \quad a_3 = -0.3554127901,$$

with k defined in (5). The numerical error in this formula for $4.2 < \beta_2 \leq 75$ is less than 0.66%, and was derived by least squares over 65 points for $\theta = 0.06(0.01)0.7$.

The modified L_L line (L_L^*) is determined by adding 2% to $\hat{\beta}_2$ for a given β_1 .

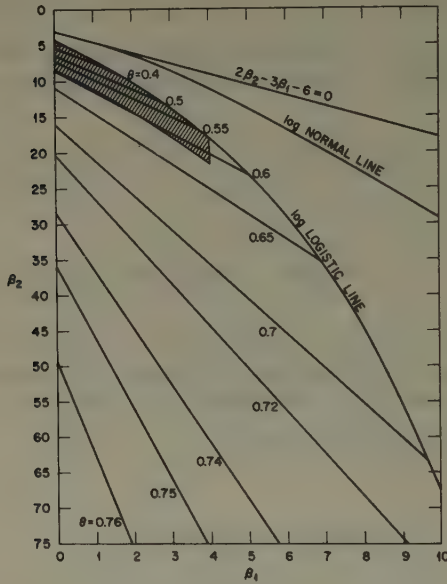


FIG. 1: A domain of validity and contours of constant θ . (The shaded area is the domain covered by the Tadikamalla and Johnson tabulation.)

5. ERROR ANALYSIS

To check accuracy, evaluate (θ_a, Ω_a) for a given couplet (β_1, β_2) in the domain of validity. Insert these solution approximates in the moment equations and evaluate feedback values β_{1f} , and β_{2f} (Table 2). It will be seen that the solutions achieve excellent accuracy.

Our assessment for the domain $\beta_1 = 0, L_L^*$, and $4.2 < \beta_2 \leq 75$ is:

TABLE 2: Examples with selected $\sqrt{\beta_1}$ and β_2 values.

β_2	$\sqrt{\beta_1}$	δ	Ω	β_2	$\sqrt{\beta_1}$	δ	Ω
5.0	0.0001	8.775311	-0.000090	25.0	1.0	4.387016	-0.290559
4.99999275	0.00010000			25.00000000	1.00000029		
5.0	0.01	8.776190	-0.008983	25.0	2.0	4.832095	-0.931755
4.99999275	0.01000000			25.00000000	2.00000014		
5.0	0.2	9.153899	-0.191669	75.0	0.1	4.085994	-0.022874
4.99999336	0.20000000			74.99999811	0.10000000		
5.0	0.25	9.393910	-0.249409	75.0	0.5	4.091768	-0.115511
5.00000000	0.25000641			75.00000000	0.49999989		
5.0	0.5	12.905460	-0.881084	75.0	1.0	4.110234	-0.238612
5.00000000	0.50000348			75.00000000	0.99999983		
25.0	0.1	4.285635	-0.026367	75.0	2.0	4.191144	-0.556868
25.00000204	0.10000000			75.00000000	2.00000009		
25.0	0.5	4.308770	-0.134789	75.0	3.0	4.357797	-1.369873
25.00000000	0.50000007			75.00000000	2.99999997		

(For specified $\beta_2, \sqrt{\beta_1}$ the approximate solution δ, Ω is shown, along with the feedback values of β_2 and $\sqrt{\beta_1}$.)

- (a) $\sqrt{\beta_1} > 0.2$ (i.e., using (7) for Ω_a)
 5 to 9 significant digits for $\sqrt{\beta_{1f}}$
 8 to 10 significant digits for β_{2f}
- (b) $0 < \sqrt{\beta_1} \leq 0.2$ (i.e., using (9) for Ω_a)
 4 correct decimal digits for $\sqrt{\beta_{1f}}$
 5 to 9 correct digits for β_{2f} .

It should be pointed out that all numerical work has been carried out in double precision arithmetic on IBM system 360 model 91, and that error assessments are related to this implementation.

The solutions provided are accurate enough for most practical situations and can be programmed on portable calculators. There will of course be a slight loss of accuracy but not serious if 10-digit input is used with scientific notation throughout.

For more restrictive computer facilities the ten-point formula in Table 1 of θ can be used followed by (7) and (9). The domain now is roughly that of the Tadikamalla and Johnson tabulation. Feedback values of $\sqrt{\beta_1}$, and β_2 can be in error by about 0.05%.

Acknowledgement. The first author's research was sponsored by the Applied Mathematical Sciences Research Program, Office of Energy Research, U. S. Department of Energy under contract W-7405-eng-26 with the Union Carbide Corporation.

REFERENCES

- Bowman, K. O. and Shenton, L. R. (1979). Evaluation of the parameters of S_U by rational fractions. *Communications in Statistics, Simulation and Computation*, B9(2), 127-132.
- Johnson, N. L. (1965). Tables to facilitate fitting S_U frequency curves. *Biometrika*, 52, 547-558.
- Tadikamalla, P. R. and Johnson, N. L. (1979). Systems of frequency curves generated by transformations of logistic variables. Technical Report, University of North Carolina.

[Received April 1980. Revised September 1980]

A METHOD FOR THE EVALUATION OF CUMULATIVE PROBABILITIES OF BIVARIATE DISTRIBUTIONS USING THE PEARSON FAMILY

RUDOLPH S. PARRISH

Computer Sciences Corporation
EPA Environmental Research Laboratory
College Station Road
Athens, Georgia 30613 USA

ROLF E. BARGMANN

Department of Statistics
University of Georgia
Athens, Georgia 30602 USA

SUMMARY. A technique is developed for the approximation of bivariate cumulative distribution function values through the use of the bivariate Pearson family when moments of the distribution to be approximated are known. The method utilizes a factorization of the joint density function into the product of a marginal density function and an associated conditional density, permitting the expression of the double integral in a form amenable to the use of specialized Gaussian-type quadrature techniques for numerical evaluation of cumulative probabilities. Such an approach requires moments of truncated Pearson distributions, for which a recurrence relation is presented, and moments of the conditional distributions for which quartic expressions are used. It is shown that results are of high precision when this technique is employed to evaluate cumulative distribution functions that are of the Pearson class.

KEY WORDS. Bivariate distribution, Pearson system, evaluation of distribution functions, quadrature.

1. INTRODUCTION

Let $f_{Y_1, Y_2}(y_1, y_2)$ represent an arbitrary joint density function (probability function in the discrete case) of the random variables Y_1 and Y_2 . It is desired to obtain approximations for values of the joint distribution function

$$F_{Y_1, Y_2}(y_1, y_2) = \int_{\ell_1}^{y_1} \int_{\ell_2}^{y_2} f_{Y_1, Y_2}(t_1, t_2) dt_2 dt_1 = \Pr\{Y_1 \leq y_1, Y_2 \leq y_2\}$$

where ℓ_1 is the lower limit for Y_1 and ℓ_2 , possibly a function of Y_1 , is the lower limit for Y_2 . Standard techniques of approximation of bivariate cumulative distribution functions include the bivariate normal approximation (Johnson, 1949) and the bivariate series of the Edgeworth or Gram-Charlier type (Ord, 1972; Mardia, 1970). In univariate distributions, the choice of the nearest member of the Pearson class leads to very satisfactory approximations (cf. White, 1960), if programs for precise evaluation are available (Bouwer & Bargmann, 1974, 1978).

If the method of approximation of the cumulative probabilities is to be generalized, a wide class of bivariate distributions should be employed. We propose to use a rather extensive family of distributions, the bivariate Pearson family (cf. Elderton & Johnson, 1969; van Uven, 1947, 1948), from which is drawn a member appropriate for the approximation. After a member of the bivariate Pearson class has been chosen which is nearest the distribution under consideration (in terms of moments up to eighth order), a technique of evaluation is employed which involves marginal and conditional distributions of the univariate Pearson class. The respective double integrals are evaluated by a specialized Gaussian quadrature technique which uses the truncated marginal distribution as kernel. Thus, all expressions can be obtained to a very high degree of precision (ten places for members of the bivariate Pearson class) utilizing merely the univariate Pearson programs.

2. MARGINAL-CONDITIONAL FACTORIZATION APPROACH

Let $f(x_1, x_2)$ represent a member of the bivariate Pearson family. This joint density may be factored as the product of $f_1(x_1)$, the marginal density for X_1 , and $f_{2|1}(x_2|x_1)$, the associated conditional density function of X_2

given that $X_1 = x_1$. Without loss of generality, assume that X_1 and X_2 are standardized with respect to location and scale.

Mardia (1970) has shown that the conditional distributions belong to the univariate Pearson family whenever the joint distribution is of the bivariate Pearson class. Although, in the general case, a similar result has not been proven for the marginal distribution (nor do there appear to be counter examples), the Pearson family contains enough joint distributions with marginals which are of the Pearson type to justify the use of that (proper or improper) subclass as an approximating family. Indeed, successful application of this procedure only requires that the marginal and conditional distributions be of such nature that univariate Pearson members provide good approximations.

The cumulative probability, $\Pr\{X_1 \leq A_1, X_2 \leq A_2\}$, may be written as

$$F(A_1, A_2) = \int_{\ell_1}^{A_1} f_1(x_1) \int_{\ell_2(x_1)}^{A_2} f_{2|1}(x_2|x_1) dx_2 dx_1. \quad (1)$$

The inner integral, being a function of x_1 and A_2 , may be denoted by $H(x_1, A_2)$ so that (1) becomes

$$F(A_1, A_2) = \int_{\ell_1}^{A_1} f_1(x_1) H(x_1, A_2) dx_1.$$

Written in this way, it is evident that the integral may be evaluated by utilizing nonstandard Gaussian quadrature (cf. Lether, 1978) treating $f_1(x_1)$ as kernel provided H can be evaluated and provided the "truncated" marginal moments,

$$m_r = E_{A_1}\{X_1^r\} = \int_{\ell_1}^{A_1} x_1^r f_1(x_1) dx_1,$$

are obtainable. [Here the constant divisor $F_1(A_1)$ is omitted.] The approximation for $F(A_1, A_2)$ is merely

$$F(A_1, A_2) \approx \sum_i W_i H(x_{1i}, A_2) \quad (2)$$

where the W_i are appropriate Gaussian weights and the x_{1i} are Gaussian points, all dependent on the truncated marginal moments.

With the marginal density function $f_1(x_1)$ serving as the kernel, we may construct a sequence of orthogonal polynomials $\{P_n\}$ where

$$P_n(x) = x^n + a_{n-1}^{(n)} x^{n-1} + \dots + a_1^{(n)} x + a_0^{(n)},$$

and

$$E_{A_1} \{P_m P_n\} = \int_{\mathcal{L}_1} P_m(x) P_n(x) f_1(x) dx = \begin{cases} 0, & \text{if } m \neq n \\ \lambda_n, & \text{if } m = n. \end{cases}$$

The Christoffel difference equation (Hamming, 1973) defines a recurrence relation useful for obtaining this sequence, it may be expressed as

$$P_{n+1}(x) = \left[x - \frac{\mu_n}{\lambda_n} - a_{n-1}^{(n)} \right] P_n(x) - \frac{\lambda_n}{\lambda_{n-1}} P_{n-1}(x) \quad (3)$$

where

$$\mu_n = E\{x^{n+1} P_n\} = m_{2n+1} + a_{n-1}^{(n)} m_{2n} + \dots + a_1^{(n)} m_{n+2} + a_0^{(n)} m_{n+1}$$

and

$$\lambda_n = m_{2n} + a_{n-1}^{(n)} m_{2n-1} + \dots + a_1^{(n)} m_{n+1} + a_0^{(n)} m_n.$$

Without loss of generality, P_0 may be taken as unity. The constant in $P_1(x) = x + a_0^{(1)}$ may be determined from the relation

$$0 = E_{A_1} \{P_0 P_1\} = \int_{\mathcal{L}_1} (x + a_0^{(1)}) f_1(x) dx = m_1 + a_0^{(1)} m_0.$$

Equation (3) may then be used to generate the rest of the sequence. In accordance with the Gaussian quadrature technique, the Gaussian points x_{1i} are the zeros of $P_N(x)$ where N is the number of points used in the summation of (2).

To determine the weights W_i , consider

$$E_{A_1} \{P_0(x) P_n(x)\} = \int_{\mathcal{L}_1} f_1(x) P_n(x) dx = \sum_{i=1}^N W_i P_n(x_{1i}),$$

for $n = 0, 1, \dots, N-1$. We obtain the following system of linear equations:

$$\sum_{i=1}^N W_i = m_0 \quad \text{and} \quad \sum_{i=1}^N W_i P_n(x_{1i}) = 0, \quad \text{for } n = 1, 2, \dots, N-1.$$

Solving this system for the weights W_i requires only simple synthetic division (Bargmann & Halfon, 1977). One aspect of this procedure is that if H were a polynomial of degree less than $2N$, then the formula (2) would be exact. Thus, given the truncated moments $E_{A_1}\{X_1^r\}$ up to order $2N-1$, the weights W_i and the points x_{1i} may be determined.

We have assumed that X_1 and X_2 are standardized random variables and that the inner integral of (1), $H(x_1, A_2)$, represents the distribution function of a Pearson distribution whose type is determined by the conditional moments, $E\{X_2^s|x_1\}$, $s = 1, 2, 3, 4$. $H(x_1, A_2)$ may be expressed in the form of the distribution function of a standardized univariate Pearson variable as

$$H(x_1, A_2) = \int_a^b f_{2|1}^*(z|x_1) dz$$

where $a = [\ell_2(x_1) - \mu_{2|1}]/\sigma_{2|1}$ and $b = [A_2 - \mu_{2|1}]/\sigma_{2|1}$,

and where $f_{2|1}^*(z|x_1)$ represents the standardized conditional density function and where $\mu_{2|1}$ and $\sigma_{2|1}$ denote the conditional mean and standard deviation, both depending explicitly on x_1 . We may further write, denoting the conditional c.d.f. by

$F_{2|1}^*$:

$$H(x_1, A_2) = F_{2|1}^* \left[\frac{A_2 - \mu_{2|1}}{\sigma_{2|1}}; \beta_{1|x_1}, \beta_{2|x_1} \right]$$

where $\beta_{1|x_1}$ and $\beta_{2|x_1}$ are the third and fourth standardized conditional moments, and thus emphasize the dependence on the first four conditional moments. Therefore, (2) becomes

$$F(A_1, A_2) \approx \sum_{i=1}^N W_i F_{2|1}^* \left[\frac{A_2 - \mu_{2|x_{1i}}}{\sigma_{2|x_{1i}}} ; \beta_{1|x_{1i}}, \beta_{1|x_{1i}} \right] \quad (4)$$

where the W_i and x_{1i} are dependent on A_1 through the truncated marginal moments.

In the next two sections we shall detail the methods for obtaining the required truncated and conditional moments using marginal and mixed moments of the distribution whose probabilities are to be approximated.

3. TRUNCATED MARGINAL MOMENTS

Let X have a Pearson distribution with probability density function $f(x)$. The truncated moments (again with the constant divisor omitted) are given by

$$E_T\{X^r\} = \int_{\ell}^T x^r f(x) dx, \text{ where } \ell \leq X.$$

When $r = 0$, $E_T\{X^r\}$ is just the cumulative distribution function $F(T)$, while for other values of r , a recurrence relation may be established to provide values for the truncated moments.

Cohen (1951) derived such a recurrence relation for doubly truncated Pearson distributions. As a special case, in a form slightly different from Cohen's the relations are

$$E_T\{X\} = \left(\frac{1}{1-2c_2} \right) [-f(T)(c_0+c_1T+c_2T^2) - (a-c_1) F(T)] \quad (5)$$

and, for $r \geq 1$,

$$E_T\{X^r\} = \left\{ \frac{1}{1-c_2(r+1)} \right\} \left[-T^r f(T)(c_0+c_1T+c_2T^2) + c_0(r-1)E_T\{X^{r-2}\} - (a-c_1r) E_T\{X^{r-1}\} \right] \quad (6)$$

where a, c_0, c_1, c_2 are the coefficients of the Pearson differential equation

$$\frac{1}{f} \frac{df}{dx} = - \frac{x+a}{c_0 + c_1x + c_2x^2}.$$

These coefficients may be expressed in terms of the moments of X (Elderton & Johnson, 1969). In the case where the mean ($E(X) \equiv \mu$) is zero, expressions for these coefficients are (Johnson & Kotz, 1970)

$$c_0 = (4\beta_2 - 3\beta_1)\mu_2/D, \quad a = c_1 = \sqrt{\beta_1}(\beta_2+3)\sqrt{\mu_2}/D,$$

and

$$c_2 = (2\beta_2 - 3\beta_1 - 6)/D, \quad (7)$$

where $D = 10\beta_2 - 12\beta_1 - 18$ and where, with $\mu_r = E\{(X-\mu)^r\}$, $\sqrt{\beta_1} = \mu_3/\mu_2^{3/2}$ and $\beta_2 = \mu_4/\mu_2^2$. If the origin is not taken at the mean, say for the random variable $Y = X + \gamma$, the truncated marginal moments may be expressed in terms of those of X as follows.

$$E_T\{\gamma^r\} = E_{T-\gamma}\{(X+\gamma)^r\} = \sum_{i=0}^r \binom{r}{i} \gamma^{r-i} E_{T-\gamma}\{X^i\}.$$

In addition, X has a Pearson distribution whose type is determined by $\sqrt{\beta_1}$ and β_2 . Hence, given moments of X to order four, the values of a , c_0 , c_1 , c_2 , $f(T)$, and $F(T)$ may all be obtained. The latter two values are available through numerical techniques and computer programs such as those by Bouver and Bargmann (1974). The recurrence relation (6) will thus provide all orders of truncated moments required for producing Gaussian points and weights for the marginal distribution $f_1(x_1)$.

4. CONDITIONAL MOMENTS

Let the joint distribution of X_1 and X_2 be of the Pearson class. Mardia (1970) has shown that the regression of X_2 on X_1 (or X_1 on X_2) is linear. This idea may be extended to show that $E\{X_2^s|x_1\}$ is of degree s in x_1 . It was noted earlier that the conditional distribution is Pearson, so that $f_{2|1}(x_2|x_1)$ satisfies a differential equation of the form

$$\frac{1}{f_{2|1}(x_2|x_1)} \frac{\partial f_{2|1}(x_2|x_1)}{\partial x_2} = \frac{L}{Q} \quad (8)$$

where L and Q are linear and quadratic functions, respectively, of x_1 and x_2 . Rearranging and multiplying (8) by x_2^s , then

integrating each side over the range of X_2 , we arrive at

$$-E\{X_2^s \frac{\partial Q}{\partial x_2} | x_1\} - s E\{X_2^{s-1} Q | x_1\} = E\{X_2^s L | x_1\} \quad (9)$$

Letting $L = a_0 + a_1 x_1 + a_2 x_2$ and $Q = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{12} x_1 x_2 + b_{22} x_2^2$ and expanding (9)

$$\begin{aligned} & -(b_2 E\{X_2^s | x_1\} + b_{12} x_1 E\{X_2^s | x_1\} + 2b_{22} E\{X_2^{s+1} | x_1\}) \\ & - s(b_0 E\{X_2^{s-1} | x_1\} + b_1 x_1 E\{X_2^{s-1} | x_1\} + b_2 E\{X_2^s | x_1\} \\ & + b_{11} x_1^2 E\{X_2^{s-1} | x_1\} + b_{12} x_1 E\{X_2^s | x_1\} + b_{22} E\{X_2^{s+1} | x_1\}) \\ & = a_0 E\{X_2^s | x_1\} + a_1 x_1 E\{X_2^s | x_1\} + a_2 E\{X_2^{s+1} | x_1\}. \end{aligned}$$

When $s = 1$, it is clear that $E\{X_2^2 | x_1\}$ is a quadratic function of x_1 . With $s = 2$, $E\{X_2^3 | x_1\}$ is a cubic function of x_1 and with $s = 3$, $E\{X_2^4 | x_1\}$ is a quartic function of x_1 . The Mardia result, i.e., that $E\{X_2 | x_1\}$ is a linear function of x_1 , is merely the special case for $s = 0$. Thus, quartic functions were chosen to provide conditional moment approximations since, if f is a member of the bivariate Pearson family, these quartic functions would be the exact moments.

The $(r,s)th$ mixed moment can be expressed using the marginal-conditional factorization regardless of whether the joint distribution is Pearson. We have

$$\begin{aligned} E\{X_1^r X_2^s\} &= \int x_1^r f_1(x_1) \int x_2^s f(x_2 | x_1) dx_2 dx_1 \\ &= \int x_1^r f_1(x_1) E\{X_2^s | x_1\} dx_1 \end{aligned}$$

where the limits of integration extend over the domain of the joint density function. Expressing $E\{X_2^s | x_1\}$ as a quartic function of x_1 , which in the case of non-Pearson distributions may represent an approximation, we have, writing μ_{rs} for $E\{X_1^r X_2^s\}$,

$$\begin{aligned} \mu_{rs} &= \int x_1^r f_1(x_1) \cdot \{w_0^{(s)} + w_1^{(s)} x_1 + w_2^{(s)} x_1^2 + w_3^{(s)} x_1^3 + w_4^{(s)} x_1^4\} dx_1 \\ &= w_0^{(s)} \mu_{r0} + w_1^{(s)} \mu_{r+1,0} + w_2^{(s)} \mu_{r+2,0} + w_3^{(s)} \mu_{r+3,0} + w_4^{(s)} \mu_{r+4,0} \end{aligned}$$

Letting $r = 0, 1, 2, 3, 4$, we obtain a persymmetric system of equations that may be solved to obtain the conditional moment weights w_i , (different from the weights W_i of (4)). This system in matrix form is

$$\begin{bmatrix} \mu_{00} & \mu_{10} & \mu_{20} & \mu_{30} & \mu_{40} \\ \mu_{10} & \mu_{20} & \mu_{30} & \mu_{40} & \mu_{50} \\ \mu_{20} & \mu_{30} & \mu_{40} & \mu_{50} & \mu_{60} \\ \mu_{30} & \mu_{40} & \mu_{50} & \mu_{60} & \mu_{70} \\ \mu_{40} & \mu_{50} & \mu_{60} & \mu_{70} & \mu_{80} \end{bmatrix} \begin{bmatrix} w_0^{(s)} \\ w_1^{(s)} \\ w_2^{(s)} \\ w_3^{(s)} \\ w_4^{(s)} \end{bmatrix} = \begin{bmatrix} \mu_{0s} \\ \mu_{1s} \\ \mu_{2s} \\ \mu_{3s} \\ \mu_{4s} \end{bmatrix}$$

or, $V \tilde{w}^{(s)} = \tilde{m}^{(s)}$. Letting $s = 1, 2, 3, 4$, we get four such matrix equations that may be combined into the single expression $V \tilde{W} = \tilde{M}$ where the (i, j) element of \tilde{W} is denoted by $w_{i-1}^{(j)}$ and that of \tilde{M} is $\mu_{i-1, j}$. So the conditional moment weights $w_i^{(s)}$ can be expressed as $\tilde{W} = V^{-1} \tilde{M}$, \tilde{W} being the solution to the linear system. The s th conditional moment approximation is simply

$$g_s(x_1) = w_0^{(s)} + w_1^{(s)} x_1 + w_2^{(s)} x_1^2 + w_3^{(s)} x_1^3 + w_4^{(s)} x_1^4. \quad (10)$$

These approximations may be used to determine the appropriate Pearson type for the conditional distribution and its limits, and, hence, permit the evaluation of the conditional cumulative distribution function, $F_{2|1}^*$, of (4).

5. ILLUSTRATIONS

Example 1: Dirichlet. The pdf of the bivariate Dirichlet distribution with parameters θ_1, θ_2 , and θ_3 is given by

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_1) \Gamma(\theta_2) \Gamma(\theta_3)} y_1^{\theta_1-1} y_2^{\theta_2-1} (1-y_1-y_2)^{\theta_3-1}$$

where $0 < y_1 < 1$, $0 < y_2 < 1 - y_1$, and $\theta_1, \theta_2, \theta_3 > 0$. This is a member of the bivariate Pearson class. Noncentral moments of the Dirichlet may be expressed in terms of ascending factorial powers as

$$\mu'_{kj} = \frac{\theta_1^{(k)} \theta_2^{(j)}}{\{\theta_1 + \theta_2 + \theta_3\}^{(k+j)}}$$

where $\theta_1^{(k)} = \theta_1(\theta_1+1)\cdots(\theta_1+k-1)$. For any given values of θ_1, θ_2 , and θ_3 , moments may be computed so that the persymmetric system (10) may be constructed. Taking, for example, $\theta_1=2$, $\theta_2=3$, and $\theta_3=1$, the matrices of equation (10) are then

$$\tilde{V} = \begin{bmatrix} 1 & 1/3 & 1/7 & 1/14 & 5/126 \\ 1/3 & 1/7 & 1/14 & 5/126 & 1/42 \\ 1/7 & 1/14 & 5/126 & 1/42 & 1/66 \\ 1/14 & 5/126 & 1/42 & 1/66 & 1/99 \\ 5/126 & 1/42 & 1/66 & 1/99 & 1/143 \end{bmatrix}$$

and

$$\tilde{M} = \begin{bmatrix} 1/2 & 2/7 & 5/28 & 5/42 \\ 1/7 & 1/14 & 5/126 & 1/42 \\ 3/56 & 1/42 & 1/84 & 1/154 \\ 1/42 & 1/105 & 1/231 & 1/462 \\ 1/84 & 1/231 & 5/2772 & 5/6006 \end{bmatrix}$$

Thus, $\tilde{W} = \tilde{V}^{-1}\tilde{M}$ is

$$\tilde{W} = \begin{bmatrix} 3/4 & 3/5 & 1/2 & 3/7 \\ -3/4 & -6/5 & -3/2 & -12/7 \\ 0 & 3/5 & 3/2 & 18/7 \\ 0 & 0 & -1/2 & -12/7 \\ 0 & 0 & 0 & 3/7 \end{bmatrix}$$

The columns of W provide the coefficients for the conditional moment approximations (11). Since the conditional distribution of $Y_2|y_1$ is, in fact a Type I Pearson member, i.e., beta, with parameters θ_2 and θ_3 and domain $0 < y_2 < 1-y_1$, the conditional moment approximations are seen to be exact.

In Section 3, the expressions (7) for c_0, c_1, c_2, a , and the recurrence formula (6) were given on the basis of $\mu'_{10} = 0$. As noted there, by letting $X_1 = Y_1 - \mu'_{10}$ and $S = T - 1/3$, the truncated marginal moments of Y_1 may be expressed as follows.

$$\begin{aligned} E_T\{Y_1^k\} &= E_{T-\mu'_{10}}\{(X_1 + \mu'_{10})^k\} \\ &= \sum_{i=0}^k \binom{k}{i} (\mu'_{10})^{k-i} E_{T-\mu'_{10}}\{X_1^i\} \\ &= \sum_{i=1}^k \binom{k}{i} \left(\frac{1}{3}\right)^{k-i} \left[\left\{ \frac{1}{1+\frac{1}{4}(i+1)} \right\} \cdot \left\{ -S^{i-1} \cdot 20T(1-T)^3 \right. \right. \\ &\quad \cdot \left[\frac{1}{18} + \frac{1}{12} S - \frac{1}{4} S^2 \right] + \frac{1}{18}(i-1)E_S\{X_1^{i-2}\} \\ &\quad \left. \left. + \frac{1}{12}(i-1)E_S\{X_1^{i-1}\} \right\} \right] + \left(\frac{1}{3}\right)^k E_S\{X_1^0\} \end{aligned}$$

where, of course, $E_S\{X_1^0\} = E_T\{Y_1^0\} = F_{Y_1}(T) = I(T; 2, 4)$, an incom-

plete beta function, here. (In the general case, this quantity is obtained as the cumulative distribution function of the appropriate Pearson member chosen on the basis of β_{10} and β_{20} .)

The marginal distribution is Pearson, beta with parameters θ_1 and $\theta_2 + \theta_3$, so the truncated marginal moment expressions are exact.

In view of the exactness of the moments, error occurring in the approximation of Dirichlet probabilities (indeed, those of any member of the Pearson class) using the marginal-conditional factorization technique would be closely related to that introduced as a result of the Gaussian quadrature routine and to the precision with which the Pearson distribution functions are evaluated. The former is largely determined by the number of

points used in the approximation. Here, for instance, $F(.7, .3)$ equals .1323 (exact) while the 3, 5, 7, 9, 11, 13, and 15-point approximations yield, respectively, 2, 4, 6, 7, 9, 11, and 13 significant digit precision. For $F(.3, .1)$, the 3-point formula yields .899988E-3 versus .0009 exact, while 11 digit accuracy is attained with only seven points, .899999999988E-3. The precision of the approximations for $F(.8, .1)$ ranges from only one digit with three points to five digits at nine points (.639997 versus .64 exact). Using the programs of Bouver and Bargmann (1974), in which ten place precision is guaranteed, the accuracy of the univariate Pearson distribution function values proved to be quite adequate in all cases considered (see Parrish, 1978). There was no indication that error in the double integral approximation was linked to evaluation of the univariate Pearson distribution functions. Indeed, for Pearson members, the precision typically obtained (eight to ten places or more) seldom required more than eleven Gaussian points.

Table 1 contains cdf values for an additional bivariate Pearson distribution, Type IVa (c.f., Johnson & Kotz) with parameters $\theta_1 = 4$ and $\theta_2 = 7$. The first entry is the approximation and the second is the exact value of the cdf for the indicated standardized values. The precision for these points varies between seven and ten places.

Example 2: Rayleigh. Consider the bivariate Rayleigh distribution with parameters n and ρ (c.f., Johnson & Kotz, 1972). Each marginal distribution is the chi-square distribution with $n-1$ degrees of freedom, thus the truncated marginal moment approximations will be exact. The distribution of $Y = S_2/(1-\rho^2)$ given $S_1 = s_1$ is the noncentral chi-square with $n-1$ degrees of freedom and noncentrality parameter $\rho^2 s_1/(1-\rho^2)$. Hence $E\{S_2^r | s_1\} = (1-\rho^2)^r \cdot E\{Y^r | s_1\}$. Further, for the standardized variables, $X_1 = (S_1 - \mu_1)/\sigma_1$, where $\mu_1 = n-1$ and $\sigma_1^2 = 2(n-1)$, the conditional moments may be expressed in terms of those of the noncentral chi-square,

$$E\{X_2^r | x_1\} = \sigma^{-r} \cdot \sum_{i=0}^r \binom{r}{i} \mu_1^i (1-\rho^2)^{r-i} (-1)^i \cdot E\{Y^{r-i} | \mu + \sigma x_1\}$$

Letting $n = 25$ and $\rho = .5$, the coefficients for computing conditional moment approximations using the marginal-conditional technique with standardized moments are given in the matrix below.

TABLE 1: Type IVa Cdf approximations (first entries) and exact values (second entries), $\theta_1 = 4, \theta_2 = 7$.

A1	A2				
	-1.8	-0.3	1.2	2.7	4.2
.5	.136952686082E-1	.391338650245	.694810441884	.732994868209	.734913857712
	.136952685984E-1	.391338650241	.694810441883	.732994868207	.734913857711
3.0		.416960249817	.880356767040	.981508433380	.989335799276
		.416960249807	.880356768048	.981508432506	.989335800374
5.5			.881535588960	.989162838293	.999253786485
			.881535588471	.989162729764	.999253791674
8.0				.989188281052	.999413142429
				.989188281173	.999415161795

$$\tilde{W} = \begin{bmatrix} 0 & 15/16 & 9\sqrt{3}/32 & 1539/512 \\ 1/4 & \sqrt{3}/16 & 99/128 & 171\sqrt{3}/256 \\ 0 & 1/16 & 3\sqrt{3}/64 & 117/256 \\ 0 & 0 & 1/64 & 3\sqrt{3}/128 \\ 0 & 0 & 0 & 1/256 \end{bmatrix}$$

For the Rayleigh distribution our approximations to cumulative probabilities were found to be consistently accurate to five or six decimal places using as few as three points, with only minor improvement observed as the number of points increased. For example, approximations for $F(-1.5, -1.5)$, $F(2,1)$ and $F(5,3)$ were, with seven points, .004301998, .8206635, and .9937327 compared to exact values of .00430254, .820658, and .993735, respectively. This should be contrasted with the rather poor precision obtained by a bivariate Edgeworth expansion (Aiuppa, 1975). Similar results were obtained for other parameter values (see Parrish, 1978). In the case of $n = 3$, $\rho = .5$, where precision might be expected to decay, the approximation provided three to six digit accuracy. Inasmuch as the moment evaluations here were shown to be exact and in light of the behavior of cumulative distribution function approximations for various values of n the error encountered must be attributed to the approximations of the conditional cumulative distribution function values. In fact, by replacing the conditional distribution function approximation with exact conditional distribution function values and using five point Gaussian quadrature, the bivariate cumulative probabilities were obtained to 12-place precision. The standardized conditional moments (Johnson & Kotz, 1970) of $S_2/(1-\rho^2)$ given $S_1 = s_1$ are

$$\sqrt{\beta_1} = \frac{\sqrt{8} \left(n-1 + \frac{3\rho^2 s_1^2}{1-\rho^2} \right)}{\left(n-1 + \frac{2\rho^2 s_1^2}{1-\rho^2} \right)^{3/2}} \quad \text{and} \quad \beta_2 = 3 + \frac{12 \left(n-1 + \frac{4\rho^2 s_1^2}{1-\rho^2} \right)}{\left(n-1 + \frac{2\rho^2 s_1^2}{1-\rho^2} \right)^2}.$$

It may be shown that $0 \leq 3\beta_2 - 4\beta_1 - 9$ and $2\beta_2 - 3\beta_1 - 6 \leq 0$, which, when given strict inequality in the latter, indicate that the Type I Pearson is the member used to provide the conditional cumulative distribution function approximations in this example. E. S. Pearson (1963) has considered the use of the Type I Pearson distribution as an approximation for the noncentral chi-square.

TABLE 2: Rayleigh Cdf approximations (first entries) and exact values (second entries),
 $n = 25, \rho = .5$.

A1	A2					
	-3.0	-1.5	0.0	1.0	2.0	3.0
-3.0	.26613E-12	.24178E-7	.11921E-6	.13878E-6	.14180E-6	.14209E-6
	.23955E-12	.24135E-7	.11891E-6	.13841E-6	.14141E-6	.14170E-6
-1.5	.25464E-7	.43020E-2	.32304E-1	.42234E-1	.44595E-1	.44940E-1
	.24135E-7	.43025E-2	.32303E-1	.42234E-1	.44595E-1	.44940E-1
0.0	.13636E-6	.32295E-1	.329792	.479786	.527611	.536962
	.11891E-6	.32303E-1	.329778	.479788	.527614	.536962
1.0	.16344E-6	.42232E-1	.479812	.730624	.820650	.840588
	.13841E-6	.42234E-1	.479788	.730623	.820658	.840588
2.0	.16840E-6	.44585E-1	.527641	.820663	.932408	.958961
	.14141E-6	.44595E-1	.527614	.820658	.932418	.958963
3.0	.16898E-6	.44930E-1	.536988	.840595	.958952	.987922
	.14170E-6	.44940E-1	.536962	.840588	.958963	.987925

Selected comparisons are presented in Table 2 for the Rayleigh distribution.

The bivariate class is in no way restricted to members in which the conditional distribution belongs to the same type of univariate Pearson class, for the different Gaussian points. Since the evaluation programs for the Pearsonian distributions select the appropriate type automatically (on the basis of the third and fourth standardized moments), this is of no concern. For non-members of the bivariate Pearson class, exact conditional moments are replaced by quartic approximations (since quartics produce exact conditional moment values for members of the Pearson class). Of interest here is whether the conditional distributions remain of the same type for all values of the marginal variate. If this presents a restriction of the bivariate Pearson class in general, then the technique described here would be utilizing a rather more extensive class since, when the approximation is performed, the conditional distribution is in no way restricted to remain in the same Pearson class type for the different Gaussian points.

REFERENCES

- Aiuppa, T. A. (1975). *Distributions of linear combinations of quadratic forms (four-moment approximations), with applications*. Ph.D. dissertation, Department of Statistics, University of Georgia.
- Bargmann, R. E. and Halfon, E. (1977). Efficient algorithms for statistical estimation in compartmental analysis: modelling ^{60}Co kinetics in an aquatic microcosm. *Ecological Modelling*, 3, 211-226.
- Bouwer, H. and Bargmann, R. E. (1974). Tables of the standardized percentage points of the Pearson system of curves in terms of β_1 and β_2 . Themis Technical Report No. 32, University of Georgia.
- Bouwer, H. and Bargmann, R. E. (1978). Optimizing the Karl Pearson coefficients in statistical curve fitting. *American Statistical Association, Proceedings of the Statistical Computing Section*, 314-319.
- Cohen, A. C. (1951). Estimation of parameters in truncated Pearson frequency distributions. *Annals of Mathematical Statistics*, 22, 256-265.
- Elderton, W. P. and Johnson, N. L. (1969). *Systems of Frequency Curves*. Cambridge University Press, London.
- Hamming, R. W. (1973). *Numerical Methods for Scientists and Engineers*. McGraw-Hill, New York.
- Johnson, N. L. (1949). Bivariate distributions based on simple translation systems. *Biometrika*, 36, 297-304.

- Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Distributions*, Vol. 1, 2. Houghton-Mifflin, Boston.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Lether, E. M. (1978). *Application of nonstandard Gaussian Quadrature to multiple integrals in statistics*. Master's thesis, Department of Statistics, University of Georgia.
- Mardia, K. V. (1970). *Families of Bivariate Distributions*. Griffin, London.
- Ord, J. K. (1972). *Families of Frequency Distributions*. Hafner, New York.
- Parrish, R. S. (1978). *Evaluation of bivariate distribution functions using the bivariate Pearson family*. Ph.D. dissertation, Department of Statistics, University of Georgia.
- Pearson, E. S. (1963). Some problems arising in approximating to probability distributions. *Biometrika*, 50, 95-111.
- van Uven, M. J. (1947, 1948). Extension of Pearson's probability distributions to two variables, I-IV. *Nederlandsche Akademie Van Wetenschappen*, 50(1947), 1063-1070, 1252-1264; 51(1958), 41-52, 191-196.
- White, J. E. (1960). Some contributions to the evaluation of Pearsonian distribution functions. NIH Technical Report No. 1, Virginia Polytechnic Institute.

[Received April 1980. Revised October 1980]

TRANSFORMATION OF A DISCRETE DISTRIBUTION TO NEAR NORMALITY

FABIÁN HERNÁNDEZ¹

University of Wisconsin

RICHARD A. JOHNSON

University of Wisconsin

SUMMARY. Utilizing an information number approach, we propose an objective method for the normalization of either discrete distributions, or sample counts, by means of a power transformation. Approximations are also given to the original known probabilities. Next, we derive the large sample distribution of our estimate of the power transformation. We compare our methods with the Box-Cox procedure, applied to observed counts, and conclude that their technique often provides good approximations even though their underlying assumption of normality is clearly violated. Two examples illustrate our methods.

KEY WORDS. Transformations, discrete distributions.

1. INTRODUCTION AND SUMMARY

The transformation or 're-expression' of counts is now common practice for the data analyst. Tukey (1977, p. 83) specifically mentions some advantages of transforming counts. Our procedure selects a 'normalizing' transformation from the family

This research was sponsored by the Office of Naval Research under Grant No. N00014-78-C-0722.

¹Now at Dirección General de Estadística, Mexico

$$X^{(\lambda)} = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(X), & \lambda = 0 \end{cases}; \quad X > 0. \quad (1)$$

considered by Box and Cox. It provides an objective way to determine λ .

In Section 2, we discuss criteria by which discrete random variables, or their transformations, may be judged to be nearly normal. Employing the Kullback-Leibler information number, we introduce a transformation technique that applies where the true underlying distribution is *known*. Also, we develop a discretization of the normal distribution that approximates a given discrete distribution.

Sample analogues, of the methods in Section 2, are developed in Section 3 and their asymptotic distribution derived. These can be used to obtain approximate confidence intervals for the 'best' transformation parameter.

For comparative purposes, we show that the Box-Cox technique leads to sensible results, provided that we add a positive constant. Section 5 includes the re-expression of counts and the normalization of raw test scores.

2. TRANSFORMATION AND APPROXIMATION OF A KNOWN DISCRETE DISTRIBUTION

2.1 Normal Approximation to Transformation of Smoothed Discrete Random Variables. Let X be a discrete random variable which takes value i with probability $p_i = P[X=i]$ for $i \geq 0$.

Then $Y = X + U$ is absolutely continuous when U is independent of X and is uniform on $[c, c+1]$ some fixed $c > 0$. Let Y have p.d.f. $g(\cdot)$, $V = Y^{(\lambda)}$ have a p.d.f. $g_\lambda(\cdot)$ and $\phi_{\mu\sigma}(\cdot)$ be the p.d.f. of a normal distribution with mean μ and standard deviation σ . In our search for a transformation, we replace the discrete variable X by the absolutely continuous Y , and then select a power transformation of Y .

Employing the Kullback-Leibler information number between g_λ and $\phi_{\mu\sigma}$, as a measure of closeness, we propose to minimize

$$I[g_\lambda; \phi_{\mu\sigma}] = E_{g_\lambda} \left\{ \log \left(\frac{g_\lambda(V)}{\phi_{\mu,\sigma}(V)} \right) \right\} \quad (2)$$

with respect to μ, σ and λ . Minimizing (2) first over μ and σ , we find that the optimal value λ_* of λ is found by minimizing

$$G(\lambda) = \frac{1}{2}[\log(2\pi) + 1] + E_g \{ \log[g(X)] \} + (1-\lambda)E_g [\log(X)] + \frac{1}{2}\log[V_g(X^{(\lambda)})] \quad (3)$$

provided that $E_g(X^{2\lambda})$ and $E_g[\log(X)]^2$ are finite. Here $V_g(X^{(\lambda)}) = E[X^{(\lambda)} - E(X^{(\lambda)})]^2$ and $E_g(\cdot)$ denotes that the expectation is taken with respect to g .

2.1.1 Proposed procedure for transforming a known discrete distribution. Replace X by the absolutely continuous random variable $Y = X+U$, where X has the given discrete distribution and U , stochastically independent of X , has a uniform distribution on the interval $[c, c+1]$, for some $c > 0$. To 'normalize' X , we make the p.d.f. of $Y^{(\lambda)}$ 'closest' (in the information number sense) to a normal p.d.f. by minimizing (2) with respect to (μ, σ, λ) .

Example 1. Let X have the *Poisson distribution* with parameter α . Let $Y = X+U$ with $U \sim U[1, 2]$ (i.e., $c=1$). The function $G(\cdot)$ defined in (3) becomes

$$\begin{aligned} G(\lambda) = & \text{const} - \lambda \sum_{i=0}^{\infty} \{ (1+i) \log[1 + 1/(1+i)] + \log(2+i) \} P_0(i, \alpha) - \lambda \\ & + \frac{1}{2} \log \left\{ \sum_{i=0}^{\infty} (1+i)^{2\lambda+1} \{ [1 + 1/(1+i)]^{2\lambda+1} - 1 \} P_0(i, \alpha) / (1+2\lambda) \right. \\ & \left. - \left\{ \sum_{i=0}^{\infty} (1+i)^{\lambda+1} \{ [1 + 1/(1+i)]^{\lambda+1} - 1 \} P_0(i, \alpha) / (1+\lambda) \right\}^2 \right\} \\ & - \log(|\lambda|) \end{aligned}$$

for $\lambda \neq -\frac{1}{2}, -1$. Here

$$\text{const} = [\log(2\pi) + 1]/2 + \alpha[\log(\alpha) - 1]$$

$$+ \sum_{i=0}^{\infty} \{ \log(i!) + (1+i)\log[2+i]/(1+i) \} + \log(2+i) \} P_0(i, \alpha)$$

$$\text{and } P_0(i, \alpha) = \alpha^i \exp(-\alpha) / i!.$$

In Figure 1, we plot $G(\lambda)$ vs λ for $\alpha = 2, 3, 4$ and 5 .

According to our criterion, the usual variance stabilizing transformation (See Bartlett, 1947, p. 41), $\lambda = \frac{1}{2}$, is a reasonable choice since it makes $|G(\lambda_*) - G(\frac{1}{2})|$ 'small' for these values of α . In Table 1, we record λ_* , the information number of the transformed variable $I[g_{\lambda_*}; \phi_{\mu_* \sigma_*}]$, the information number of the untransformed variable $I[g_1; \phi_{\mu_V \sigma_V}]$, $\mu_V = \alpha + \frac{1}{2}$, and $\sigma_V^2 = \alpha + 1/12$. We also include the information number $I[g_{\frac{1}{2}}; \phi_{\mu_*(\frac{1}{2}) \sigma_*(\frac{1}{2})}]$ corresponding to the square root transformation.

2.2 Normal Approximation of Discrete Probabilities. Alternatively we can approximate the probability p_i by a probability q_i obtained from a normal c.d.f. To measure the accuracy of this type of approximation we utilize the Kullback-Leibler information number in its discrete population version (see Kullback, 1968, p. 128).

Let $d \in (0, 1)$. We propose to approximate p_i by

$$q_i \equiv q_i(\theta) = \Phi\left[\frac{(i+d)^{(\lambda)} - \mu}{\sigma}\right] - \Phi\left[\frac{(i-d)^{(\lambda)} - \mu}{\sigma}\right], \quad i \geq 1. \quad (4)$$

where $\Phi(\cdot)$ is the c.d.f. of a standard normal distribution.

Here $q_0 = \Phi\left[\frac{(1-d)^{(\lambda)} - \mu}{\sigma}\right]$ and if $p_i = 0$ for $i > N$,

$$q_N = 1 - \Phi\left[\frac{(N-1+d)^{(\lambda)} - \mu}{\sigma}\right]. \quad \text{Let } P = \{p_i : i \geq 0\} \text{ and}$$

$Q = \{q_i : i \geq 0\}$ with q_i defined in (4). We often take $d = .5$.

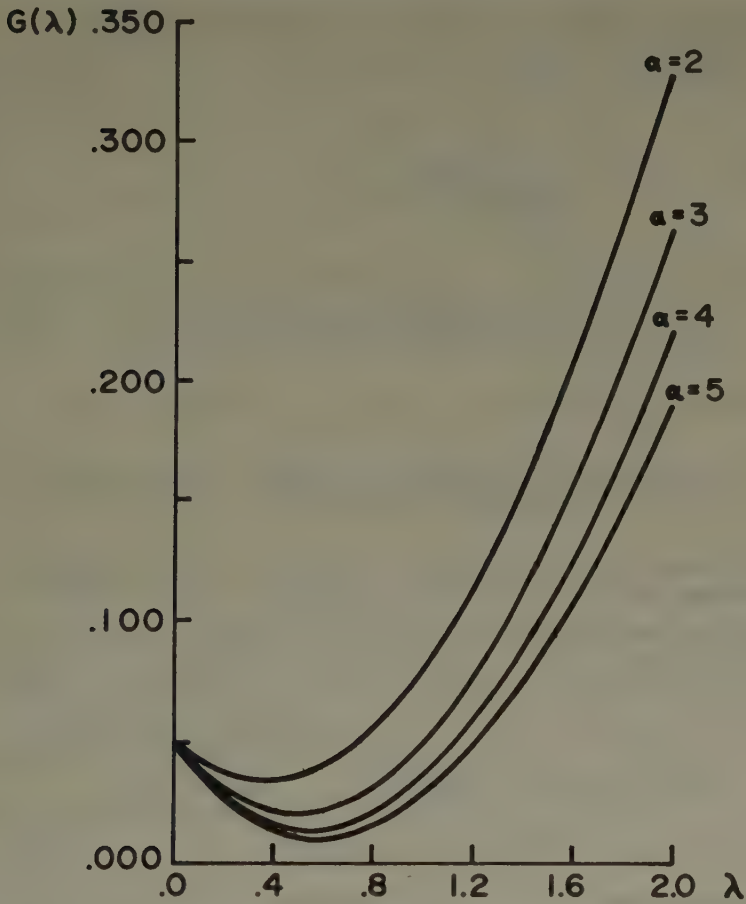


Fig. 1: The function $G(\cdot)$ for the Poisson distribution.

TABLE 1: Transformation and comparison of information numbers for the Poisson distribution.

α	λ_*	$I[g_{\lambda_*}; \phi_{\mu_* \sigma_*}]$	$I[g_1; \phi_{\mu_v \sigma_v}]$	$I[g_{1/2}; \phi_{\mu_*(1/2) \sigma_*(1/2)}]$
2	.36889	.033815	.08104	.035925
3	.48816	.019620	.05047	.019641
4	.54318	.013090	.03572	.013308
5	.57540	.009729	.02753	.013167

2.2.2 *Proposed procedure for the normal approximation of probabilities.* Let $d \in (0,1)$ be fixed. Minimize the Kullback-Leibler information number

$$I[P;Q] = p_0 \log \left\{ \frac{p_0}{\phi\left[\frac{(1-d)^{(\lambda)} - \mu}{\sigma}\right]} \right\} + \sum_{i=1}^{\infty} p_i \log \left\{ \frac{p_i}{\phi\left[\frac{(i+d)^{(\lambda)} - \mu}{\sigma}\right] - \phi\left[\frac{(i-d)^{(\lambda)} - \mu}{\sigma}\right]} \right\} \quad (5)$$

with respect to $\theta' = (\mu, \sigma, \lambda)$.

3. TRANSFORMATION OF COUNTS AND NORMAL APPROXIMATION OF OBSERVED PROPORTIONS

3.1 *Transformation of Counts.* Let X be a discrete random variable taking the value i with probability $p_i = P[X=i] > 0$, for $i = 0, 1, \dots, N < \infty$. Let X_1, \dots, X_n be i.i.d. as X and denote the frequency of the value i by f_{in} . Set $I_{\{i\}} =$ the indicator of $[X_j=i]$ so $f_{in} = \sum_{j=1}^n I_{\{i\}}(X_j)$ is the frequency of $[X=i]$ and the relative frequency is

$$\hat{p}_{in} = f_{in}/n \text{ for } i = 0, 1, \dots, N. \quad (6)$$

Once having observed x_1, x_2, \dots, x_n , we treat these as possible values of a random variable Y_n , where Y_n now takes the value i with probability \hat{p}_{in} . Next, construct

$$V_n = Y_n + U_n, \text{ with p.d.f. } g_n(\cdot) \quad (7)$$

where Y_n and U_n are stochastically independent and U_n has a uniform distribution on the interval $[\beta, \beta+1)$, for some fixed $\beta > 0$. Let $W_n = V_n^{(\lambda)}$ have p.d.f. $g_{n\lambda}(\cdot)$ and $\phi_{\mu\sigma}(\cdot)$ be the p.d.f. of a normal distribution with mean μ and standard deviation σ .

Now, suppose we want to transform the data so that they appear to come from a normal distribution. We propose to transform V_n , instead of the original observations x_i , $i = 1, \dots, n$

and to select a transformation λ , in such a way that the distribution of $V_n^{(\lambda)}$ is 'closest' to a normal distribution in the Kullback-Leibler information number sense. That is, we minimize

$$I[g_{n\lambda}; \phi_{\mu\sigma}] = E_{g_{n\lambda}} \left\{ \log \left[\frac{g_n(W_n)}{\phi_{\mu\sigma}(W_n)} \right] \right\} \quad (8)$$

with respect to μ , σ and λ . Thus, minimizing first with respect to μ and σ , it is easily shown that the optimal value, λ_{*n} , of λ is obtained by minimizing

$$G_n(\lambda) = \frac{1}{2} [\log(2\pi) + E_{g_n} \{\log[g_n(V_n)]\} + (1-\lambda) E_{g_n} [\log(V_n)] + \frac{1}{2} \log[V_{g_n}(V_n^{(\lambda)})]]. \quad (9)$$

In summary,

3.1.1 Proposed procedure for transforming counts to near normality. Having observed x_1, x_2, \dots, x_n , introduce the discrete random variable Y_n which takes the value i with probability $\hat{p}_{in} = \frac{1}{n} \sum_{j=1}^n I_{\{i\}}(x_j)$. Replace Y_n by the absolutely continuous random variable $V_n = Y_n + U_n$, where U_n is independent of Y_n and is uniform on the interval $[\beta, \beta+1)$, some fixed $\beta > 0$. In order to 'normalize' the observations, we transform V_n employing Procedure 2.1.1.

The transformation λ is selected by minimizing (9) with respect to λ .

The asymptotic behavior of $\theta_{*n} = (\mu_{*n}, \sigma_{*n}, \lambda_{*n})'$, the value of θ that minimizes (8), is given in the following result.

Theorem 3.1. Let X , be a discrete distribution with $p_i = P[X=i]$, $i \leq N$, and $W = (X + U)^{(\lambda)}$ have p.d.f. $g_\lambda(\cdot)$ where U is uniform and independent of X . Set $\theta' = (\theta_1, \theta_2, \theta_3) = (\mu, \sigma, \lambda)$ and suppose that the following conditions are satisfied:

1) The parameter space Θ is a compact set given by

$$\Theta = \{\theta = (\mu, \sigma, \lambda)' \mid |\mu| \leq M, c \leq \sigma \leq d, a \leq \lambda \leq b; -\infty < a < 0 < b, c, d, M < \infty\}.$$

ii) $H(\tilde{\theta}) = I[g_\lambda; \phi_{\mu\sigma}]$ has a unique minimum at

$$\tilde{\theta}_* = (\mu_*, \sigma_*, \lambda_*)' \in \Theta.$$

Then,

1) $\lim_{n \rightarrow \infty} \tilde{\theta}_{*n} = \tilde{\theta}_*$, with probability one.

If we further assume that

iii) $\tilde{\theta}_*$ is an interior point of Θ and

iv) $\nabla^2 H(\tilde{\theta}_*) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} H(\tilde{\theta}) \right) \Big|_{\tilde{\theta}=\tilde{\theta}_*}$, is non-singular,

Then

2) $\sqrt{n}(\tilde{\theta}_{*n} - \tilde{\theta}_*) \xrightarrow{d} N_3(0, VWV')$ where $V = [\nabla^2 H(\tilde{\theta}_*)]^{-1}$ and

$W = [E[C_k(X_1, \tilde{\theta}) \ C_\ell(X_1, \tilde{\theta})]]$ with

$$C_1(X_1, \tilde{\theta}) = - \sum_{i=0}^N \int_{i+\beta}^{i+\beta+1} \left(\frac{v^{(\lambda)} - \mu}{\sigma} \right)^2 dv \ I_{\{i\}}(X_1),$$

$$C_2(X_1, \tilde{\theta}) = \frac{1}{\sigma} - \frac{1}{\sigma} \sum_{i=0}^N \int_{i+\beta}^{i+\beta+1} \left(\frac{v^{(\lambda)} - \mu}{\sigma} \right)^2 dv \ I_{\{i\}}(X_1)$$

$$C_3(X_1, \tilde{\theta}) = \sum_{i=0}^N \int_{i+\beta}^{i+\beta+1} \left[\left(\frac{v^{(\lambda)} - \mu}{\sigma^2} \right) \frac{\partial v^{(\lambda)}}{\partial \lambda} - \log(v) \right] dv \ I_{\{i\}}(X_1).$$

Proof. See Hernandez and Johnson (1979).

Remark. Theorem 3.1 says that $\tilde{\theta}_{*n}$ converges, with probability one, to $\tilde{\theta}_*$ the value of $\tilde{\theta}$ that minimizes the Kullback-Leibler information number between g_λ , the p.d.f. of $(Y+U)^{(\lambda)}$, and a normal p.d.f. Hence, Procedure 2.1.1 can be interpreted as the infinite-sample analogue of the current technique.

3.2 A Normal Approximation to Observed Proportions. We want to approximate observed proportions (6) by a set of normal probabilities $Q(\tilde{\theta}) = \{q_i(\tilde{\theta}) : 0 \leq i \leq N\}$ given by (4).

3.2.1 Proposed procedure to approximate observed proportions. In order to approximate $\hat{p}_n = \{\hat{p}_{in} : 0 \leq i \leq N\}$, by a collection

of normal probabilities (4) we minimize the Kullback-Leibler information number

$$I[\hat{p}_n ; Q(\theta)] = \sum_{i=0}^N \hat{p}_{in} \log \frac{\hat{p}_{in}}{q_i(\theta)} \tag{10}$$

with respect to θ .

Theorem 3.2. Let X assume the value i with probability p_i for $i = 0, 1, \dots, N$ where N is finite. Let X_1, \dots, X_n be i.i.d. with the same distribution as X and \hat{p}_{in} be the observed proportion of the value i . Set $\theta' = (\theta_1, \theta_2, \theta_3) = (\mu, \sigma, \lambda)$ and assume that the following conditions are satisfied.

- i) The parameter space Θ is as in Theorem 3.1,
- ii) $F(\theta) = I[P; Q(\theta)] = \sum_{i=0}^N p_i \log \frac{p_i}{q_i(\theta)}$ has a unique minimum at θ_0 .

Then, $\dot{\theta}_n$, the value which minimizes (10), satisfies

- 1) $\lim_{n \rightarrow \infty} \dot{\theta}_n = \theta_0$, with probability one.

If we further assume that

- iii) θ_0 is an interior point of Θ and
- iv) $V^2 F(\theta_0) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} F(\theta) \right) \Big|_{\theta = \theta_0}$, is non-singular,

then

- 2) $\sqrt{n}(\dot{\theta}_n - \theta_0) \xrightarrow{d} N_3(0, VWV')$ where $V = [V^2 F(\theta_0)]^{-1}$ and

$$W = \left[\sum_{i=0}^N \frac{\partial \log q_i(\theta)}{\partial \theta_k} \cdot \frac{\partial \log q_i(\theta)}{\partial \theta_\ell} p_i \right].$$

Proof. See Hernandez and Johnson (1979).

4. A COMPARISON WITH THE BOX-COX PROCEDURE APPLIED TO DISCRETE OBSERVATIONS

In our model $[X=0]$ may have positive probability so, we consider $Y = X + c$ with $c > 0.5$. The Box and Cox (1964) method selects λ by maximizing

$$\begin{aligned} \ell_{\max}(\lambda) = & -\frac{n}{2}[\log(2\pi)+1] - \frac{n}{2} \log\left\{ \sum_{i=1}^n [Y_i^{(\lambda)} - \overline{Y^{(\lambda)}}]^2 / n \right\} \\ & + (\lambda-1) \sum_{i=1}^n \log(Y_i) \end{aligned} \quad (11)$$

with respect to λ .

Let $Y_* = X + U$ where U is independent of X and is uniform on $[c-\frac{1}{2}, c+\frac{1}{2}]$. Procedure 2.1.1, the large sample limit of Procedure 3.1.1, selects λ to minimize $G(\lambda)$. Asymptotically the Box-Cox approach, applied directly to the discrete observations, requires the minimization of a function $-\phi(\lambda)$ and

$$G(\lambda) = -\phi(\lambda) + \text{constant} + \text{Error}$$

where

$$\begin{aligned} \text{Error} = & \lambda \sum_{i=0}^N p_i \int_{i+c-\frac{1}{2}}^{i+c+\frac{1}{2}} [\log(i+c) - \log(y)] dy \\ & - \frac{1}{2} \log \left[\frac{\sum_{i=0}^N p_i (i+c)^{2\lambda} - \left[\sum_{i=0}^N p_i (i+c)^{\lambda} \right]^2}{\left[\sum_{i=0}^N p_i \int_{i+c-\frac{1}{2}}^{i+c+\frac{1}{2}} y^{2\lambda} dy - \left[\sum_{i=0}^N p_i \int_{i+c-\frac{1}{2}}^{i+c+\frac{1}{2}} y^{\lambda} dy \right]^2 \right]} \right] \end{aligned} \quad (12)$$

$$\text{When } \left| \int_{i+c-\frac{1}{2}}^{i+c+\frac{1}{2}} [\log(y) - \log(i+c)] dy \right| \leq \frac{1}{2} \frac{1}{i+c-\frac{1}{2}} \quad (13)$$

and, for $r = 1, 2$

$$\left| \int_{i+c-\frac{1}{2}}^{i+c+\frac{1}{2}} [y^{r\lambda} - (i+c)^{r\lambda}] dy \right| \leq \frac{|r\lambda|}{2} (i+c-\frac{1}{2})^{r\lambda-1}, \quad r\lambda < 1 \quad (14)$$

are small, we would expect the Error to be small. Consequently, the Box-Cox procedure and Procedure (3.1.1) should give nearly the same answer when the sample size is large.

From (13) and (14) making c large appears to improve the agreement between the two procedures. The most frequently employed transformations $\sqrt{X+c}$, $\sqrt[4]{X+c}$, $\log(X+c)$ and $(X+c)^{-1}$ satisfy $r\lambda \leq 1$.

5. APPLICATION TO THE RE-EXPRESSION OF COUNTS AND NORMALIZATION OF TEST SCORES

Example 2. Tukey (1977, p. 572), displays the following counts for the duration (in days) of incubation for 1663 eggs of the ridley turtle

days	50	51	52	53	54	55	56	57	58	59	60	61	//	64	65
number of eggs	77	122	10	321	725	180	162	21	14	6	9	14	.	1	1

He suggests the re-expression $\sqrt{\text{days}-49}$.

Setting $\beta = 0$, we apply the Proposed Procedure 3.1.1 to the above data and obtain $\theta_{*,1663} = (3.570, 1.333, 0.815)'$. Moreover, using the limiting distribution of θ_{*n} , derived in Theorem 3.1, we can establish an approximate confidence interval for λ_* . The estimated standard error of $\lambda_{*,1663}$ is $\sqrt{6.015/1663}$. Hence, an approximate 95% confidence interval for λ_* is (0.697, 0.933). Notice that $\lambda = \frac{1}{2}$ is not included in the interval.

Example 3. Ghiselli (1964, p. 78), proposes the use of the square root transformation for the normalization of the 100 test scores. We set $\beta = 0$. The application of the Proposed Procedure 3.1.1 yields $\theta_{*,100} = (4.66, 0.54, 0.070)$. Utilizing the limiting distribution of θ_{*n} , we obtain the estimated standard error of $\lambda_{*,100}$ and then approximate 95% confidence interval (-0.30, 0.44) for λ_* . Figure 2 presents a comparison of the relative frequency histograms of a) the transformed scores and b) the original scores. We also applied the Box-Cox procedure to the above scores. The estimated power transformation is $\hat{\lambda} = 0.078$ which is in good agreement with the value $\lambda_{*,100} = 0.070$.

The sample procedures introduced in Section 3, can also be applied to situations where the observations can only be ordered.

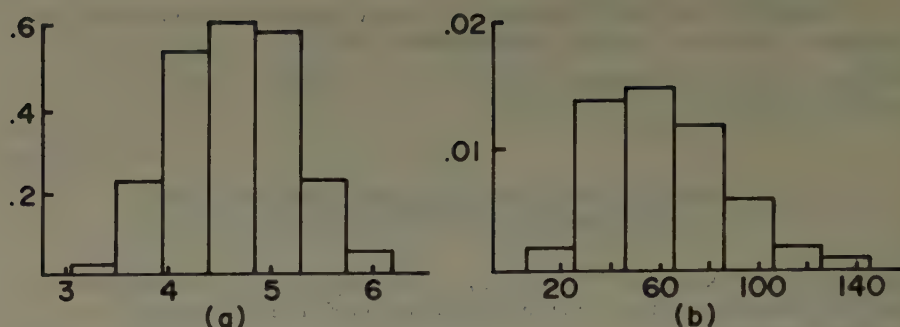


FIG. 2: Relative frequency histogram of (a) the transformed scores and (b) the original scores.

That is, the observations can be assigned to exactly one of the categories $1, 2, \dots, N$. Here N corresponds to the highest category, $N-1$ to the second highest, etc. With this scoring, we are able to apply our methods to the relative frequencies of the categories.

REFERENCES

- Bartlett, M. S. (1947). The use of transformations. *Biometrics*, 3, 39-52.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-243; discussion, 244-252.
- Ghiselli, E. E. (1964). *Theory of Psychological Measurement*. McGraw-Hill, New York.
- Hernandez, F. and Johnson, R. A. (1979). Transformation of a discrete distribution to near normality. Technical Report No. 546, Department of Statistics, University of Wisconsin.
- Kullback, S. (1968). *Information Theory and Statistics*. Dover, New York.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.

[Received May 1980. Revised October 1980]

MULTIVARIATE DISTRIBUTIONS IN RELIABILITY THEORY AND LIFE TESTING

HENRY W. BLOCK

University of São Paulo and University of Pittsburgh

THOMAS H. SAVITS

University of Pittsburgh

SUMMARY. Multivariate parametric distributions which are of interest in reliability theory and life testing are discussed. These include distributions with exponential, Weibull and gamma univariate marginal distributions. Other distributions of interest are the multivariate nonparametric distributions whose marginals have increasing failure rates (IFR), increasing failure rate averages (IFRA), are new better than used (NBU) or new better than used in expectation (NBUE). Also mentioned are univariate and multivariate processes which have associated with them distributions in the various nonparametric classes mentioned above.

KEY WORDS. multivariate exponential distributions, multivariate Weibull distributions, multivariate gamma distributions, multivariate exponential extensions, shock models, threshold model, gestation model, characteristic function equation, multivariate IFR, multivariate IFRA, multivariate NBU, multivariate NBUE.

1. INTRODUCTION

In this paper we discuss the various multivariate parametric and nonparametric classes of distributions which are of interest in reliability theory and life testing.

In Section 2 we concentrate on parametric distributions whose univariate marginals are either exponential, Weibull or gamma. We also discuss multivariate exponential extensions which

are distributions whose marginals are not generally exponential but which were formulated utilizing concepts based on univariate exponential distributions.

Multivariate nonparametric classes are discussed in Section 3. These classes contain multivariate distributions which have increasing failure rate or increasing failure rate average or are new better than used or new better than used in expectation. Many formulations have been given for each of these classes, so that in this section only the most recent or the most established formulations are discussed in detail.

The paper concludes with a discussion of univariate and multivariate stochastic processes which are related to the nonparametric classes to which Section 3 was devoted.

2. PARAMETRIC DISTRIBUTIONS

2.1 Introduction. The univariate parametric distributions which have been most useful in reliability theory have been the exponential and Weibull distributions. Others which have been of some importance are the lognormal and the gamma distributions. A history of the use of these distributions is given in Chapter 1 of Barlow and Proschan (1965). An introduction to these distributions, models from which they arise, their properties and their use in reliability theory are contained in Mann, Shafer and Singpurwalla (1974), and in Barlow and Proschan (1965, 1975). See also Bain (1978). For more detailed expositions on these distributions and comprehensive bibliographies see Johnson and Kotz (1972).

Multivariate parametric distributions which are analogs of the univariate distributions previously mentioned are still being developed. Unlike the multivariate normal distribution, for most other multivariate distributions having marginals of one type there are many possible dependence structures and consequently many multivariate versions. Several multivariate exponential and related distributions, for example, have been developed. Many of these, along with their properties, are given in Basu and Block (1975) and in Block (1975). See also Johnson and Kotz (1972).

2.2 Bivariate Exponential and Related Distributions. Most of the distributions treated in this section have multivariate analogs, but for the purposes of clarity and exposition we will treat the bivariate case first. As with all the distributions discussed we will say a distribution is a multivariate "----" if all of its univariate marginals are "----". Therefore a multivariate exponential distribution will be one whose univariate

marginals are exponential. We will say a distribution is a *multivariate exponential extension* if it is derived from properties of a univariate exponential. The univariate marginals of such a distribution will not necessarily be exponential.

Freund distribution. Freund (1961) introduced one of the first bivariate distributions based on a model involving the exponential distribution. An interpretation of this model, given in Block (1975), is the following. Consider a two component system where the failure of one component affects the lifetime of the other component. Let Y_1 and Y_2 have independent exponential distributions with means α_1^{-1} and α_2^{-1} respectively (the initial distributions of the unaffected component life-times). Let Y'_1 and Y'_2 be independent of Y_1 and Y_2 and have independent exponential distributions with means $(\alpha'_1)^{-1}$ and $(\alpha'_2)^{-1}$ (the distributions of the affected components). Then it is easily shown that the distribution of the component lifetimes

$$(X_1, X_2) = \begin{cases} (Y_1, Y_1 + Y'_2) & \text{if } Y_1 < Y_2, \\ (Y_2 + Y'_1, Y_2) & \text{if } Y_2 < Y_1, \end{cases}$$

has the survival function

$$\bar{F}(x_1, x_2) = \begin{cases} \frac{\alpha_1}{\alpha_1 + \alpha_2 - \alpha'_2} \exp[-(\alpha_1 + \alpha_2 - \alpha'_2)x_1 - \alpha'_2 x_2] \\ + \frac{\alpha_2 - \alpha'_2}{\alpha_1 + \alpha_2 - \alpha'_2} \exp[-(\alpha_1 + \alpha_2)x_2] & \text{if } x_1 \leq x_2, \\ \frac{\alpha_2}{\alpha_1 + \alpha_2 - \alpha'_1} \exp[-\alpha'_1 x_1 - (\alpha_1 + \alpha_2 - \alpha'_1)x_2] \\ + \frac{\alpha_1 - \alpha'_1}{\alpha_1 + \alpha_2 - \alpha'_1} \exp[-(\alpha_1 + \alpha_2)x_1] & \text{if } x_1 > x_2. \end{cases} \quad (1)$$

Properties of this distribution are given in Johnson and Kotz (1972).

Marshall and Olkin and a related distribution. The distribution of Marshall and Olkin (1967a) has survival function

$$\begin{aligned}\bar{F}(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) \\ &= \exp[-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)] \\ &\quad \text{for } t_1 > 0, t_2 > 0,\end{aligned}$$

where $\lambda_1, \lambda_2, \lambda_{12}$ are nonnegative. This distribution is derivable from 1) a fatal shock model, 2) a nonfatal shock model, and 3) a loss of memory model. For details on these models and for properties of this important distribution see Johnson and Kotz (1972) and Barlow and Proschan (1975). Estimation and testing for this model have been carried out in Bennis, Bain and Higgins (1972), Bhattacharya and Johnson (1973) and Proschan and Sullo (1975). A more general version of this distribution is given by Marshall and Olkin (1967b).

A recent characterization of this distribution due to Block (1977a) is that (T_1, T_2) has this distribution if and only if T_1 and T_2 are marginally exponential and $\min(T_1, T_2)$ is exponential and independent of $T_1 - T_2$. Other characterizations of this distribution are given in Galambos and Kotz (1978) and in Basu and Block (1975).

A distribution which is closely related to the Marshall and Olkin distribution has been studied by Block and Basu (1974). It is also closely related to the Freund distribution and can be obtained from the interpretation of Freund's model given previously where the effect of one component on the other is a strain, i.e., $\alpha_1 < \alpha'_1, \alpha_2 < \alpha'_2$. This distribution is obtained from (1) with the choices

$$\alpha_i = \lambda_i + \lambda_{12} \lambda_i / (\lambda_1 + \lambda_2), \quad \alpha'_i = \lambda_i + \lambda_{12},$$

where $\lambda_1, \lambda_2, \lambda_{12}$ are nonnegative and $\lambda_1 + \lambda_2 > 0$. It turns out that this distribution is also derivable from a loss of memory model similar to that of Marshall and Olkin. Furthermore this distribution is the absolutely continuous part of the Marshall and Olkin distribution. It should also be noted that the lifetime of the two-organ system of Gross, Clark and Liu (1971) and also of the two-organ subsystem of Gross (1973) is a special case of the maximum lifetime of this distribution. See the original paper for details, estimation and other properties.

Estimation and testing have also been carried out by Mehrotra and Michalek (1976) and by Gross and Lam (1979). The latter authors consider an application of this distribution to bivariate relapse times for patients receiving different treatments.

Friday and Patil distribution. Proschan and Sullo (1974) briefly suggest a model which incorporates both the Marshall and Olkin and the Freund distributions. Friday and Patil (1977) pursue the idea of a distribution containing these two distributions still further. They develop a similar although more general distribution than that of Proschan and Sullo which is derivable from (1) a threshold model, (2) a gestation model and (3) a warmup model. The survival function can be written as

$$\bar{F}(x_1, x_2) = \alpha_0 \bar{F}_A(x_1, x_2) + (1 - \alpha_0) \bar{F}_S(x_1, x_2) \quad (2)$$

where $\bar{F}_A(x_1, x_2)$ is given by (1), i.e., the Freund distribution, and

$$\bar{F}_S(x_1, x_2) = \exp[-(\alpha_1 + \alpha_2) \max(x_1, x_2)] \quad \text{if } x_1 > 0, x_2 > 0.$$

It is clear that for $\alpha_0 = 1$ equation (2) gives the Freund distribution. It should be noticed that this equation is of the form of the Marshall and Olkin distribution (see Theorem 1.5, Chapter 5 of Barlow and Proschan (1975)). The choice of parameters $\alpha_0 = (\lambda_1 + \lambda_2) / (\lambda_1 + \lambda_2 + \lambda_{12})$, where λ_1 , λ_2 and λ_{12} are positive and α_1 , α_1' , α_2 , α_2' as chosen for the Block and Basu distribution, yields the Marshall and Olkin distribution.

Also contained in the paper of Friday and Patil is a nice summary of the relations among the various distributions just discussed. Also discussed are transformations from independence for the distribution and computer generation and efficiency for this distribution.

Downton distribution. This distribution is a special case of a classical bivariate gamma distribution due to Wicksell and to Kibble. See Krishnaiah and Rao (1961) for a discussion of this gamma distribution and references. Downton (1970) developed a model which gave rise to this bivariate exponential distribution and proposed its use in the setting of reliability theory.

An interpretation of this model, which is due to Arnold (1975b) is presented here. Consider a two component system where the components are each subjected to nonfatal shocks which occur according to two independent Poisson processes. Assume the processes have rates $(1-\rho)/\mu_1$ and $(1-\rho)/\mu_2$ respectively where

$0 < \rho < 1$, $0 < \mu_1$, $0 < \mu_2$. Let $\{X_{ij} : i = 1, \dots, \infty\}$, $j = 1, 2$, represent interarrival times for each of the two processes. Assume that each component fails after N_1 and N_2 shocks respectively where $N_1 \equiv N_2 \equiv N$ is a geometric random variable with parameter $1-\rho$ (i.e., $P\{N=n\} = \rho^{n-1}(1-\rho)$, $n = 1, 2, \dots$). The times to failure of the two components are then given by

$$(Y_1, Y_2) = \left(\sum_{i=1}^{N_1} X_{i1}, \sum_{i=1}^{N_2} X_{i2} \right). \quad (3)$$

By conditioning on N and using characteristic functions it is easily seen that (Y_1, Y_2) has density

$$f(y_1, y_2) = \frac{\mu_1 \mu_2}{1-\rho} \exp\left(-\frac{\mu_1 y_1 + \mu_2 y_2}{1-\rho}\right) I_0\left(\frac{2\sqrt{\rho \mu_1 \mu_2 y_1 y_2}}{1-\rho}\right) \quad (4)$$

for $y_1 > 0$, $y_2 > 0$, where I_0 is the modified Bessel function of the first kind of order 0. This is the bivariate exponential distribution given by (2.10) of Downton (1970).

Downton also shows that if instead of $N_1 \equiv N_2 \equiv N$ in his derivation (and equivalently in the above derivation) we let (N_1, N_2) assume various other bivariate geometric distributions a distribution of the same form as (4) is obtained.

As mentioned initially the above distribution can be obtained as the special case of a particular bivariate gamma distribution. This distribution is obtained as follows. Let (X_1, Y_1) , $(X_2, Y_2), \dots, (X_n, Y_n)$ be iid bivariate standard normal distributions with correlation ρ . Then

$$\left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2 \right)$$

has a correlated bivariate gamma (chi-square) distribution. For the case $n=2$, a distribution of the form (4) is obtained.

Hawkes distribution. The bivariate exponential distribution of Hawkes (1972) is obtained from the same model as that of Downton. The only difference is in the choice of the bivariate geometric distribution (N_1, N_2) . This bivariate geometric distribution

which was derived by Hawkes was derived independently by Esary and Marshall (1973) and by Arnold (1975a). The model given here is from the Esary and Marshall paper.

Consider two devices which receive nonfatal shocks at discrete time periods labeled by the positive integers. The occurrence of these shocks is independent and at each cycle there is a probability p_{11} that both devices survive the shocks, probability p_{10} that the first survives but the second does not, probability p_{01} that the second survives and the first does not and p_{00} that both devices fail. Then letting N_i be the number of shocks to failure for devices $i = 1, 2$, and conditioning on the occurrence of the first cycle, as in Arnold (1975b), it is easily seen that the characteristic function $\phi(t_1, t_2)$ satisfies

$$\phi(t_1, t_2) = e^{it_1 + it_2} [p_{00} + p_{01}\phi(0, t_2) + p_{10}\phi(t_1, 0) + p_{11}\phi(t_1, t_2)]. \quad (5)$$

This is essentially the characteristic function equation of Paulson and Uppuluri (1972b) and is easily solved (i.e. take $t_1 = 0$ and solve for $\phi(0, t_2)$, then for $\phi(t_1, 0)$ similarly, then for $\phi(t_1, t_2)$). The bivariate distribution which has this characteristic function is given by

$$P(N_1 > n_1, N_2 > n_2) = \begin{cases} p_{11}^{n_1} (p_{01} + p_{11})^{n_2 - n_1} & \text{if } n_1 \leq n_2 \\ p_{11}^{n_2} (p_{10} + p_{11})^{n_1 - n_2} & \text{if } n_2 \leq n_1 \end{cases} \quad (6)$$

Using this distribution in the model of Downton, Hawkes in slightly different notation (i.e., $p_{ij} = p_{ji}$ for all $i = 0, 1, j = 0, 1$) obtains his distribution. By taking $p_{01} = p_{10} = 0$ it can be seen that $N_1 \equiv N_2$ and so the Downton distribution is a special case of Hawkes. The resulting transform is given in Hawkes (1972) along with some properties.

Paulson distribution. Paulson (1973) derives a bivariate exponential distribution through a characteristic function equation. This equation is the generalization of a one-dimensional characteristic function equation which arises from a compartment model (see Paulson and Uppuluri, 1972a). A generalization of the compartment model also leads to the bivariate equation.

The bivariate equation is given by

$$\phi(t_1, t_2) = \psi(t_1, t_2) [p_{00} + p_{01}\phi(0, t_2) + p_{10}\phi(t_1, 0) + p_{11}\phi(t_1, t_2)]$$

where $p_{00} + p_{01} + p_{10} + p_{11} = 1$, $p_{10} + p_{11} < 1$, $p_{01} + p_{11} < 1$ and $\psi(t_1, t_2) = [(1 - i\theta_1 t_1)(1 - i\theta_2 t_2)]^{-1}$. Then solving for $\phi(t_1, t_2)$ in the above equation leads to the bivariate characteristic function

$$\begin{aligned} \phi(t_1, t_2) = & [(1 - i\theta_1 t_1)(1 - i\theta_2 t_2) - p_{11}]^{-1} [p_{00} \\ & + p_{10}(1 - i\mu_1 t_1)^{-1} + p_{01}(1 - i\mu_2 t_2)^{-1}] \end{aligned}$$

where $\mu_1 = \theta_1(p_{00} + p_{01})^{-1}$ and $\mu_2 = \theta_2(p_{00} + p_{10})^{-1}$.

It can be shown that this is exactly the form of the Hawkes distribution and henceforth we will refer to this distribution as the Hawkes-Paulson distribution. For properties see Paulson (1973) and Hawkes (1972).

Arnold classes. In describing these classes Arnold uses what he calls a generalized multivariate geometric distribution. It is easily seen that this is a reparametrized version of (6) in the bivariate case. Thus we let (N_1, N_2) be the bivariate distribution given by (6). Then Arnold's bivariate classes $\epsilon_n^{(2)}$ consist of the random variables

$$(Y_1, Y_2) = \left(\sum_{i=1}^{N_1} X_{i1}, \sum_{i=1}^{N_2} X_{i2} \right)$$

where (X_{i1}, X_{i2}) for $i = 1, 2, \dots$ are bivariate iid rvs with distributions in $\epsilon_{n-1}^{(2)}$ for $n > 1$ and where $\epsilon_0^{(2)}$ consists of (X, X) where X is exponential. Clearly the marginals are exponential for all the classes. It also is not hard to show that $\epsilon_1^{(2)}$ contains the pair of independent exponentials (see Arnold, 1975a) and also contains the Marshall and Olkin distribution (see the nonfatal shock model discussed in Barlow and Proschan, 1975). Furthermore, it follows from the derivations given here of the Downton and the Hawkes distributions that these are contained in $\epsilon_2^{(2)}$ since $\epsilon_1^{(2)}$ contains the independent exponentials.

The Arnold classes of distributions have been described using the characteristic function equation approach of Paulson and Uppuluri (1972b) and Paulson (1973) in Block, Paulson and Kohberger (1975). In this latter paper the characteristic function equation approach has been used to derive properties of the distributions in this class, including descriptions of the standard distributions in the class, infinite divisibility of the distributions, moment properties and asymptotic properties. These results are summarized, without proof, in Block (1977b), in which it is also shown how the distributions in the class lead to multivariate shock models of the type studied in the univariate case by Esary, Marshall, and Proschan (1973).

2.3 Multivariate Exponential and Related Distributions. Most of the bivariate models in the preceding section have multivariate ($n > 3$) analogs. In general the ideas are similar to the bivariate case, but the notational complexity is greatly increased. Without giving many details, we will briefly discuss the multivariate situation.

The Freund distribution has been generalized to the multivariate case by Weinman (1966) but only for identically distributed marginals. See Johnson and Kotz (1972) for details concerning this distribution. Block (1975) has considered a generalization of the Freund distribution for the case when the marginals need not be identically distributed as well as generalizing the Block and Basu (1974) and the Proschan and Sullo (1974) models in the same paper.

Generalization of the Downton (1970), Hawkes (1972) and Paulson (1973) distributions implicitly exist within the framework of the general multivariate gamma distribution of Krishnamoorthy and Parthasarathy (1951) (see also Krishnaiah and Rao, 1961; Krishnaiah, 1977) and also within the framework of the Arnold classes. A specific parametric form has been given in Hsu, Shaw and Tyan (1977).

Recently a multivariate exponential distribution has been proposed by Bryant (1979) which arises in the context of certain cycling systems.

2.4 Multivariate Weibull Distributions. Multivariate Weibull distributions were discussed by Marshall and Olkin (1967a) in the context of their discussion on multivariate exponential distributions. Specifically they define a multivariate Weibull distribution by assuming (T_1, \dots, T_n) has their multivariate exponential distribution and then considering

$$(T'_1, \dots, T'_n) = (T_1^{1/\alpha_1}, T_2^{1/\alpha_2}, \dots, T_n^{1/\alpha_n}) \quad (7)$$

where $\alpha_i > 0$ for $i = 1, \dots, n$ which then has univariate Weibull marginal distributions. This procedure certainly could be extended for any multivariate exponential distribution. Moeschberger (1974) has studied bivariate Weibull distributions of this form, deriving properties and discussing maximum likelihood expectation.

David (1974) and Lee and Thompson (1974) have introduced multivariate Weibull distributions of the form (T_1, \dots, T_n) where $T_i = \min(U_j : i \text{ in } J), \phi \neq J \subset \{1, \dots, n\}, P\{U_j > x\} = \exp\{-\lambda_j x^{\alpha_j}\}, x > 0$ and U_j are independent. These distributions need not have Weibull marginals if the α_j are not all equal. Arnold (1967) has also considered Weibull distributions of a similar form, but his restriction that they belong to an additive family forces $\alpha_j = \alpha$ for all J . Thus these distributions are also of the form (7).

Recently Spurrier and Weier (1979) have modified the Freund model using Weibull instead of exponential distributions.

3. NONPARAMETRIC CLASSES OF DISTRIBUTIONS

Various classes of distributions which describe the way in which component lifetimes wear out have been discussed by many authors. The most important of these classes are 1) the increasing failure rate (IFR) class, 2) the increasing failure rate average (IFRA) class, 3) the new better than used (NBU) class and 4) the new better than used in expectation (NBUE) class. These have been extensively discussed in the literature. The case where the lifetimes are independent (which we call the *univariate case*) are discussed in the book of Barlow and Proschan (1975) and also in the expository paper of Block and Savits (1980). The case where the components are dependent (called the *multivariate case*) has also been discussed in the latter paper, but since the development in the field is so rapid many new results have appeared since this last mentioned paper. We give a brief introduction to the univariate case, some background in the multivariate case and then outline the most recent developments.

3.1 Univariate Classes. The most prominent of the nonparametric classes used in reliability theory is the class of distribution which have increasing failure rate. See Block and Savits (1980) for background and motivation for this class. In the following we let T be a random variable with distribution function $F(x)$ such that $F(0) = 0$ (the usual assumption is

$F(0^-) = 0$, but for the purpose of exposition we use the above) having density $f(x)$ (if it exists). We say F has *increasing failure rate* (IFR) if the survival function $\bar{F}(x) = P\{T > x\}$ satisfies

$$\bar{F}(x+t)/\bar{F}(x) \text{ decreases in } x > 0 \text{ for } t > 0 \quad (8)$$

or equivalently (if the density exists)

$$r(x) = f(x)/\bar{F}(x) \text{ increases in } x > 0.$$

A distribution F has *increasing failure rate average* (IFRA) if

$$\bar{F}(\alpha t) \geq \bar{F}^\alpha(t) \text{ for all } 0 < \alpha \leq 1 \text{ and all } t \geq 0$$

or equivalently (if the density exists)

$$t^{-1} \left(\int_0^t r(x) dx \right) \text{ increases in } t > 0.$$

Another equivalent formulation (see Block and Savits, 1976) is

$$\int_0^\infty h^\alpha(x/\alpha) dF(x) \geq \left\{ \int_0^\infty h(x) dF(x) \right\}^\alpha \text{ for all } 0 < \alpha \leq 1 \quad (9)$$

and all nonnegative increasing functions h .

A distribution F is *new better than used* (NBU) if

$$\bar{F}(x+t) \leq \bar{F}(x)\bar{F}(t) \text{ for all } x \geq 0, t \geq 0 \quad (10)$$

and is *new better than used in expectation* (NBUE) if

$$\int_t^\infty \bar{F}(x) dx \leq \mu \bar{F}(t) \text{ for all } t > 0$$

where $\mu = \int_0^\infty \bar{F}(x) dx$ is finite.

Dual versions for all the above definitions exist by reversing the monotonicity or the inequality. Since the treatment of these concepts is similar we shall omit it.

3.2 Multivariate Classes.

As in the parametric case many multivariate extensions are possible. Many versions of multivariate IFR, IFRA, NBU and NBUE have been proposed. For various IFR and IFRA extensions see Marshall (1974) and Esary and Marshall (1979) respectively. In the following we discuss the particular multivariate IFR and IFRA notations which at this time appear to be the most important ones. Various multivariate concepts of NBU and NBUE are given in Block

and Savits (1980). As of yet, no clear favorites have emerged but there have been several recent papers on this subject. We shall attempt to describe some of this development. In the following we let F be the multivariate distribution of the random variable $\tilde{T} = (T_1, \dots, T_n)$ which is assumed to satisfy $\bar{F}(0) = 1$.

The distribution F is said to be *MIFR* if

$$\bar{F}(\tilde{x} + t\tilde{1})/\bar{F}(\tilde{x}) \text{ decreases in } \tilde{x} \geq \tilde{0} \text{ for all } t > 0 \quad (11)$$

where $\tilde{1} = (1, \dots, 1)$.

This generalizes (8) and has many important properties (see Block and Savits, 1980). Various variants of this condition are possible (see Marshall, 1979) but the above version best captures the intuitive idea of the model that components in the same environment run for the same time (i.e., $t_1 = t_2 = \dots = t_n = t$) but may be of different ages (i.e., $\tilde{x} = (x_1, \dots, x_n)$). This concept also satisfies many of the important basic properties that one would expect of such a multivariate generalization. See Chapter 5 of Barlow and Proschan (1975) for the statements and proofs of these properties.

The concept of multivariate IFRA which we now discuss is a generalization of (9). We say F is *MIFRA* if

$$E[h(\tilde{T})] \leq E[h^\alpha(\tilde{T}/\alpha)] \text{ for all } 0 < \alpha \leq 1$$

and for all continuous nonnegative increasing functions h . Recall that F is the df of \tilde{T} . A distribution satisfying this condition has all of the properties one would expect of a generalization of the univariate IFRA concept. See Block and Savits (1979b).

Esary and Marshall (1979) have proposed various other concepts of multivariate IFRA, many of them having intuitive appeal. Unfortunately all of them fail to satisfy at least one of the basic properties which the MIFRA distributions possess. This is demonstrated in Block and Savits (1978b).

Block and Savits (1980) describe a wide variety of possible definitions for both the concepts of multivariate NBU and multivariate NBUE. Some of these were definitions given by Buchanan and Singpurwalla (1977), others were based on the multivariate IFRA concepts of Esary and Marshall and still others were based on Laplace transform characterizations of NBU and NBUE which appeared in Block and Savits (1979a) and on other

characterizations of NBU and NBUE which appeared in Block and Savits (1978a). At the time the paper of Block and Savits (1980) was written the only thing that was clear was that many multivariate NBU and NBUE concepts were possible. Since then some order has begun to appear in this field.

Marshall and Shaked (1979) introduce a compelling concept of NBU. This definition is that a random vector $\tilde{T} = (T_1, \dots, T_n)$ is multivariate NBU if

$$P\{\tilde{T} \in (\alpha + \beta)A\} \leq P\{\tilde{T} \in \alpha A\}P\{\tilde{T} \in \beta A\} \text{ for all } \alpha > 0, \beta > 0 \quad (12)$$

and all upper (or increasing) sets A in R^n . (A is an upper set if $\tilde{x} \in A$ and $\tilde{x} \leq \tilde{y}$ imply $\tilde{y} \in A$). It is clear that this is a general version of the type of definition studied by Buchanan and Singpurwalla (1977) and very recently by Ghosh and Ebrahimi (1980), i.e.,

$$\bar{F}(\tilde{x} + \tilde{y}) \leq \bar{F}(\tilde{x})\bar{F}(\tilde{y}) \text{ for all } \tilde{x} > \tilde{0} \text{ and } \tilde{y} > \tilde{0} \quad (13)$$

where \tilde{x} and \tilde{y} are perhaps further constrained. Furthermore Marshall and Shaked have four alternate characterizations of (12). Several of these involve the concept that \tilde{T} is multivariate NBU if and only if $g(\tilde{T})$ is univariate NBU for all g of a certain type. Many other properties are proven.

The classes of multivariate NBU and NBUE introduced by Buchanan and Singpurwalla (1977), where the NBU distributions are defined by properties which are cases of (13) and the NBUE distributions are integrated versions of these, have been further studied by Ghosh and Ebrahimi (1980). These authors study the relationships among these definitions (and some variants of them), their properties and also demonstrate how some multivariate shock models give rise to them. They also make connections with some of the IFRA concepts. Griffith (1979) has also considered multivariate shock models leading to some of these concepts.

A recent paper by El-Newehi, Proschan and Sethuraman (1980) discuss the multivariate class which arises as minimums of independent univariate NBU random variables. These distributions arise in the same way as the Marshall and Olkin distribution and also in the same way as one of the definitions of multivariate IFRA (i.e., Condition C) of Esary and Marshall (1979). These have properties similar to those of distributions with exponential minimums studied by Esary and Marshall (1979). Various relationships and properties are given. One of them is given that

$\tilde{T} = (T_1, \dots, T_n)$ has this NBU property then it has the Marshall and Olkin distribution if and only if, for example, $\min_{1 \leq i \leq n} T_i$ is exponential. Relationships are given between the present definition and various other definitions.

3.3 Processes. The concept of a one dimensional stochastic process being of a type described by one of the four classes has been proposed by Ross (1979). Essentially he has discussed processes which are decreasing (or increasing) and whose first entry times into a state are all IFRA. For these processes, which he calls IFRA processes, he proves a closure theorem. He also studies NBU processes. NBU processes are also considered by El-Newehi, Proschan and Sethuraman (1978) who also prove a closure theorem.

Extensions of Ross's ideas to multivariate processes have been accomplished by Block and Savits (1979c). These authors study several types of multivariate processes having IFRA type properties.

Recently Arjas (1979) has considered IFRA processes. He has discussed both the univariate and multivariate cases.

ACKNOWLEDGEMENTS

Research has been partially supported by ONR Contract N00014-76-C-0839. Research of the first named author has also been partially supported by FINEP and FAPESP while at the University of São Paulo, São Paulo, Brazil. Research of the second named author has also been partially supported by NSF Grant MCS77-01458.

REFERENCES

- Arjas, E. (1979). A stochastic process approach to multivariate reliability systems. Unpublished report.
- Arnold, B. C. (1967). A note on multivariate distributions with specified marginals. *Journal of American Statistical Association*, 62, 1460-1461.
- Arnold, B. C. (1975a). A characterization of the exponential distribution by multivariate geometric compounding. *Sankhya, Series A*, 37, 164-173.
- Arnold, B. C. (1975b). Multivariate exponential distributions based on hierarchical successive damage. *Journal of Applied Probability*, 12, 142-147.
- Bain, L. J. (1978). *Statistical Analysis of Reliability and Life-Testing Models*. Marcel Dekker, New York.

- Barlow, R. E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York.
- Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life-Testing: Probability Models*. Holt, Rinehart and Winston, New York.
- Basu, A. P. and Block, H. (1975). On characterizing univariate and multivariate exponential distributions with applications. In *Statistical Distributions in Scientific Work*, Vol. 3, G. P. Patil, S. Kotz and J. K. Ord, eds. Reidel, Dordrecht-Holland.
- Bemis, B. M., Bains, L. J. and Higgins, J. J. (1972). Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. *Journal of American Statistical Association*, 67, 927-929.
- Bhattacharya, G.K. and Johnson, R. A. (1973). On a test of independence in a bivariate exponential distribution. *Journal of American Statistical Association*, 68, 704-706.
- Block, H. W. (1975). Continuous multivariate exponential extensions. In *Reliability and Fault Tree Analysis*, R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds. SIAM, Philadelphia.
- Block, H. W. (1977a). A characterization of a bivariate exponential distribution. *Annals of Statistics*, 5, 808-812.
- Block, H. W. (1977b). A family of bivariate life distributions. In *The Theory and Applications of Reliability*, Vol. I, C. P. Tsokos and I. Shimi, eds. Academic Press, New York.
- Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *Journal of American Statistical Association*, 69, 1031-1037.
- Block, H. W., Paulson, A. S. and Kohberger (1976). Some bivariate exponential distributions: syntheses and properties. Unpublished report.
- Block, H. W. and Savits, T. H. (1976). The IFRA closure problem. *Annals of Probability*, 4, 1030-1032.
- Block, H. W. and Savits, T. H. (1978a). Shock models with NBUE survival. *Journal of Applied Probability*, 15, 621-628.
- Block, H. W. and Savits, T. H. (1978b). The class of MIFRA lifetimes and its relation to other classes. University of Pittsburgh Research Report #78-01.
- Block, H. W. and Savits, T. H. (1979a). Laplace transforms for classes of life distributions. *Annals of Probability*, to appear.
- Block, H. W. and Savits, T. H. (1979b). Multivariate IFRA distributions. *Annals of Probability*, to appear.
- Block, H. W. and Savits, T. H. (1979c). Multivariate IFRA processes. *Annals of Probability*, to appear.
- Block, H. W. and Savits, T. H. (1980). Multivariate classes in reliability theory. *Mathematics of Operations Research*, to appear.

- Bryant, J. L. (1979). Stochastic properties of cycling systems with random lifetimes. Unpublished report.
- Buchanan, W. B. and Singpurwalla, N. D. (1977). Some stochastic characterizations of multivariate survival. In *The Theory and Applications of Reliability, Vol. I*, C. P. Tsokos and I. Shimi, eds. Academic Press, New York.
- David, H. A. (1974). Parametric approaches to the theory of competing risks. In *Reliability and Biometry*, F. Proschan and R. J. Serfling, eds. SIAM, Philadelphia.
- Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society, Series B*, 32, 408-417.
- El-Newehi, E., Proschan, F. and Sethuraman, J. (1978). Multistate coherent systems. *Journal of Applied Probability*, 15, 675-688.
- El-Newehi, E., Proschan, F. and Sethuraman, J. (1980). A multivariate new better than used class derived from a shock model. FSU Statistics Report M540.
- Esary, J. D. and Marshall, A. W. (1973). Multivariate geometric distributions generated by a cumulative damage process. Naval Postgraduate School Report NP555EY73041A.
- Esary, J. D. and Marshall, A. W. (1974). Multivariate distributions with exponential minimums. *Annals of Statistics*, 2, 84-96.
- Esary, J. D. and Marshall, A. W. (1979). Multivariate distributions with increasing hazard rate average. *Annals of Probability*, 7, 359-370.
- Freund, J. (1961). A bivariate extension of the exponential distribution. *Journal of American Statistical Association*, 56, 971-977.
- Friday, D. S. and Patil, G. P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variables. In *Theory and Applications of Reliability, Vol. I*, C. P. Tsokos and I. Shimi, eds. Academic Press, New York.
- Galambos, J. and Kotz, S. (1978). *Characterizations of Probability Distributions*. Springer-Verlag, Berlin.
- Ghosh, M. and Ebrahimi, N. (1980). Multivariate NBU and NBUE Distributions. Unpublished report.
- Griffith, W. (1980). Multistate reliability models. *Journal of Applied Probability*, to appear.
- Gross, A. J. (1973). A competing risk model: a one organ subsystem plus a two organ subsystem. *IEEE Transactions on Reliability*, R-22, 24-27.
- Gross, A. J., Clark, V. A. and Liu, V. (1971). Estimation of survival parameters when one of two organs must function for survival. *Biometrics*, 27, 369-377.
- Gross, A. J. and Lam, C. F. (1979). Paired observations from a survival distribution. Unpublished report.

- Hawkes, A. G. (1972). A bivariate exponential distribution with applications to reliability. *Journal of the Royal Statistical Society, Series B.*, 34, 129-131.
- Hsu, C. L., Shaw, L. and Tyan, S. G. (1977). Reliability applications of multivariate exponential distributions. Polytechnic Institute of New York Report POLY-EE-77-036.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Krishnaiah, P. R. (1977). On generalized multivariate gamma type distributions and their applications in reliability. In *Theory and Applications of Reliability, Vol. I*, C. P. Tsokos and I. Shimi, eds. Academic Press, New York.
- Krishnaiah, P. R. and Rao, M. M. (1961). Remarks on a multivariate gamma distribution. *American Mathematical Monthly*, 68, 342-346.
- Krishnamoorthy, A. S. and Parthasarathy, M. (1951). A multivariate gamma type distribution. *Annals of Mathematical Statistics*, 22, 549-557.
- Lee, L. and Thompson, Jr., W. A. (1974). Results on failure time and pattern for the series system. In *Reliability and Biometry*. F. Proschan and R. J. Serfling, eds. SIAM, Philadelphia.
- Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.
- Marshall, A. W. (1975). Multivariate distributions with monotone hazard rate. In *Reliability and Fault Tree Analysis*, R. E. Barlow, J. Fussell and N. D. Singpurwalla, eds. SIAM, Philadelphia.
- Marshall, A. W. and Olkin, I. (1967a). A multivariate exponential distribution. *Journal of American Statistical Association*, 62, 30-44.
- Marshall, A. W. and Olkin, I. (1967b). A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4, 291-302.
- Marshall, A. W. and Shaked, M. (1979). A class of multivariate new better than used distributions. Unpublished report.
- Mehrotra, K. G. and Michalek, J. E. (1976). Estimation of parameters and tests of independence in a continuous bivariate exponential distribution. Unpublished manuscript.
- Moeschberger, M. L. (1974). Life tests under dependent competing causes of failure. *Technometrics*, 16, 39-47.
- Paulson, A. S. (1973). A characterization of the exponential distribution and a bivariate exponential distribution. *Sankhya, Series A*, 35, 69-78.
- Paulson, A. S. and Uppuluri, V. R. R. (1972a). Limit laws of a sequence determined by a random difference equation governing a one compartment system. *Mathematical Biosciences*, 13, 325-333.

- Paulson, A. S. and Uppuluri, V. R. R. (1972b). A characterization of the geometric distribution and a bivariate - geometric distribution. *Sankhya, Series A.*, 34, 88-91.
- Proschan, F. and Sullo, P. (1974). Estimating the parameters of a bivariate exponential distribution in several sampling situations. In *Reliability and Biometry*, F. Proschan and R. J. Serfling, eds. SIAM, Philadelphia.
- Proschan, F. and Sullo, P. (1976). Estimating the parameters of a multivariate exponential distribution. *Journal of American Statistical Association*, 71, 465-472.
- Ross, S. (1979). Multivalued state component systems. *Annals of Probability*, 7. 379-383.
- Spurrier, J. D. and Weier, D. R. (1979). A bivariate survival model derived from a Weibull distribution. University of South Carolina Statistics Technical Report No. 46.

[Received July 1980. Revised October 1980]

HAZARD RATES BASED ON ISOPROBABILITY CONTOURS

I. R. GOODMAN

Naval Research Laboratory
Washington, D. C. 20375 USA

SAMUEL KOTZ

University of Maryland
College Park, Maryland 20742 USA

SUMMARY. A new "scalar" definition of a (contour) multivariate hazard rate based on variation of probability distribution across the isoprobability contours is motivated and described. Several examples involving specific multivariate distributions justifying the usefulness of this definition are presented. General structures of "constant" contour multivariate hazard rates as well as increasing contour multivariate hazard rates are described.

KEY WORDS. hazard rates, isoprobability contours, logarithmic transform, power transform, multivariate distributions, exponential distribution, characterizations.

1. INTRODUCTION

Numerous devices and algorithms for generating multivariate distributions have been developed over the past few decades. Some of these are briefly sketched below without commenting on their relative merits.

1) Perhaps the earliest method was to formally generalize a system or equation defining a univariate distribution to the

multivariate case. The papers extending Pearson's system of distributions, e.g., van Uven (1947), Sagrista (1952), and Styen (1960), can be cited as examples of this approach.

2) The most direct approach - modelling - is presented in the works of Freund (1961) and Marshall and Olkin (1967a,b), whereby a multivariate distribution is obtained by indicating a specific stochastic model comprising a system of several "components" representing the random variables under consideration. The multinomial distribution is a primary motivation for such a construction.

3) Another method defines a multivariate distribution by explicitly specifying the mathematical relations between the joint distribution and its marginals. This was initiated by Fréchet (1951) and extended by Morgenstern (1956), Gumbel (1960), Farlie (1960), and Johnson and Kotz (1975a), among others.

4) Still another approach postulates a specified multivariate form for the density by reproducing in the multivariate setting the functional form of a particular univariate family of densities. This includes linear-exponential type distributions originated by Bildikar and Patil (1968), quadratic-exponential type by Day (1969) and multivariate θ -generalized distributions of Goodman and Kotz (1973).

5) Recently, Higgins (1975) has shown that given any "reasonable" set of contours generated by the relation $t(V) = y$, $V \in \mathbb{R}^n$, y real, and any continuous univariate density g , there exists a (unique) multidimensional density h with corresponding r.v., say \bar{X} , such that the density of $t(\bar{X})$ is g and the isoprobability contours of h are determined by $t(V) = y$. Thus, a transformation is displayed which maps pairs (g, t) into a large class of multivariate densities. Higgins' procedure yields a variety of multivariate distributions. However, in this "semi-modelling" approach, it is necessary to extract from the resulting multivariate distribution, the possible hidden dependencies between the component variables. This task is seldom easy, but a promising avenue is to trace the relation between the joint hazard rates corresponding to the multivariate distribution and the rates of the marginal components (see, e.g., Johnson and Kotz, 1975b).

A related method to determine the interrelation among the component variables of the multivariate distribution is to investigate how certain characteristics of a distribution, in addition to the density itself, change when a univariate distribution is extended to a multivariate setting. One such characteristic is the "hazard rate". This problem is investigated in the present paper.

2. MOTIVATION OF THE DEFINITION

Based on Higgins' (1975) results and our previous generalizations of univariate generalized normal distributions using contour concepts (Goodman and Kotz, 1973), we shall present a new definition of hazard rate for both a univariate and multivariate setting.

The "ordinary" univariate hazard rate for a random variable (r.v.) with density f and distribution function F is defined as

$$H(t) = \frac{-d \log(1-F(t))}{dt} = \frac{f(t)}{1-F(t)},$$

for all values of t for which the denominator is positive. Conversely, $1-F(t) = \exp\{-D_t^{-1} H(t)\}$, $t \geq 0$.

Multivariate generalizations of hazard rates have been proposed by Harris (1970), Basu (1971), Brindley and Thompson (1972), Block (1973, 1977), Johnson and Kotz (1975b), and Marshall (1975), among others. The last three generalizations are based on vector-valued definitions.

Here, we base our univariate as well as multivariate definitions of hazard rate on isoprobability contours (or sets).

First we present a result closely related to Higgins' original theorem (1975). We shall obtain Higgins' original result with a uniqueness property and a partial converse to his theorem (see Remark below).

Theorem 1. Let n be any fixed positive integer. Let $t: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ be any nonsingular function, i.e., $dt(\underline{x})/d\underline{x} \neq 0$ for all $\underline{x} \in D$, such that $V_{t,y} = \text{vol}\{\underline{x} | \underline{x} \in D \text{ and } t(\underline{x}) \leq y\}$ is finite, nowhere constant in y and everywhere differentiable in y , for each y in range (t) . (We assume range (t) to be some finite or infinite interval.) Let g be any continuous univariate nowhere-zero density over range (t) . Then there exists a unique multivariate density $h = \phi(t)$, corresponding to r.v. \underline{Y} , say, such that: (1) the set of isoprobability contours of h is the same as that of t , i.e., of the form $\{\underline{x} | t(\underline{x}) = \text{constant}\}$; (2) the r.v. $t(\underline{Y})$ has the univariate density g ; and (3) the specific relation between h , t , and g is

$$h(\underline{x}) = \phi(t(\underline{x})) = g(t(\underline{x})) / (D_y V_{t,y}) \Big|_{y=t(\underline{x})}. \quad (1)$$

Remark. Let h be a multivariate density over $D \subseteq R^n$ with the same properties of those of function t stated in Theorem 1.

Let t be a monotonic function, ϕ^{-1} , of h . Then: (1) the set of isoprobability contours of h is the same as that of t and (2) the r.v. $t(U)$ has density g determined by the following integrals (where U is a r.v. possessing density h):

$$g(y) = \int \frac{h(\underline{s})}{\left| \frac{dt(\underline{s})}{d\underline{s}} \right|} d\underline{s} = \phi(y) \cdot \frac{dv_{t,y}}{dy}, \quad (2)$$

where the integration is over the set $\{\underline{s} | t(\underline{s}) = y\}$.

Sketch of Proofs. The essence of Theorem 1 is Higgins' result. Uniqueness follows from the additional requirement that $V_{t,y}$ is nowhere constant (equivalently $D_y(V_{t,y}) > 0$). The Remark follows from equation (2) using the result

$$\frac{dv_{t,y}}{dy} = \int \frac{1}{\left| \frac{dt(\underline{s})}{d\underline{s}} \right|} d\underline{s}$$

with the integration over $\{\underline{s} | t(\underline{s}) = y\}$.

We now define, for a given multivariate density h , the concept of a hazard function. According to Theorem 1, a pair (g, t) uniquely defines h . However, given h , we have in general an infinite class of corresponding pairs (g, t) which are all compatible (in the sense of Theorem 1) with the given h . In order to eliminate the ambiguity, we must choose a particular t , which in turn uniquely determines g . The problem of non-uniqueness of the density g (via different choices of $t(\underline{x})$), is not unlike the one encountered in Bayesian inference when different choices of the prior distribution lead to different posterior distributions and, consequently, to estimators with different structural properties. A similar situation holds in connection with ancillary statistics (see, e.g., Fu, 1974). The devices used in Bayesian inference for choosing a natural (or "conjugate") prior compatible with the distribution at hand can readily be adapted in our case.

Basically, we are dealing with four types of multivariate distributions:

1) Those of the general exponential type, which includes linear exponential families (Bildikar and Patil, 1968), θ -generalized multivariate distributions (Goodman and Kotz, 1973), and various exponential multivariate distributions, e.g., Gumbel (1960), Freund (1961), Marshall and Olkin (1967). The natural choice for $t(\underline{x})$ for this family is the monotonically decreasing logarithmic transformation $t(\underline{x}) = -\log(h(\underline{x})/c)$, where c is an appropriate constant - affecting only the location and scale of the resulting univariate density g . This transformation is meaningful in the sense that it yields the *information content* (or the log-likelihood) of the random sample $\underline{X} = (X_1, X_2, \dots, X_n)$ distributed with density $h(\underline{x})$. The basic function of a transform is to assure the finiteness of the volume $V_{t,y}$ in

Theorem 1. (A conceptually analogous device is used by Fu, 1974.)

2) The second class of multivariate distributions is of the "power-type" form which includes multivariate t , multivariate Pareto, multivariate Burr, etc. These distributions are also defined on an infinite domain. Here, the natural choice for $t(\underline{x})$ is the monotonically decreasing power transform which does not distort the functional form of the multivariate density h , as far as the variable and the parameters are concerned, but simplifies the form of the constant part related to the location and scale parameters. This transform also does not affect the information contained in the sample $\underline{X} = (X_1, \dots, X_n)$ of the multidimensional variable with the p.d.f. $h(\underline{x})$, and serves to guarantee the finiteness of the volume $V_{t,y}$.

3) The third class of multivariate distributions has the form $x^\alpha \exp(-\beta x^\alpha)$ defined on positive orthants, as exemplified by the multivariate gamma distributions. This case can also be dealt with efficiently using information-content-type monotonic decreasing transformations $t(\underline{x}) = -\log(h(\underline{x})/c)$. In these cases, however, closed expressions for the ordinary hazard rate as well as for the contour hazards are not usually available, although the functions are computable using numerical techniques.

4) Finally, we have distributions on a finite domain corresponding to the Beta-type distribution in the univariate case, and to Dirichlet-type distributions in the multivariate case. These distributions do not require any transforms to assume the finiteness of the volume and the natural choice here is $t(\underline{x}) = h(\underline{x})/c$. As in case 3), closed expressions are not always available but both hazard rates are computable.

We shall now present the definition and several examples of distributions with the corresponding contour hazard rates and compare them, in univariate cases, with the "ordinary" rates.

3. THE BASIC DEFINITION

Let h be any given multivariate (or univariate) density over some appropriate domain $D \subseteq \mathbb{R}^n$ subject to the restrictions of Theorem 1, with corresponding r.v. \underline{U} . Let t be a monotonic (increasing or decreasing, depending on the form of h) function, ϕ^{-1} , of h . Let the density of $t(\underline{U})$ be denoted by g . Via Theorem 1 we have the relations

$$g(y) = \int \frac{h(\underline{s})}{\left| \frac{dt(\underline{s})}{d\underline{s}} \right|} d\underline{s} = \phi(y) \cdot \int \frac{1}{\frac{dt(\underline{s})}{d\underline{s}}} d\underline{s} = \phi(y) \left(\frac{dV_{t,y}}{dy} \right).$$

with the integration over $\{\underline{s} | t(\underline{s}) = y\}$. Denoting by G the cumulative distribution function corresponding to g , we define the (relative) contour hazard rate function Q as the "ordinary" hazard rate of the univariate density g , i.e., $Q(y) = g(y)/(1-G(y))$.

If the random variable with the density $h(\underline{x})$ is restricted to $D \subseteq \mathbb{R}^{n+}$ and represents joint life-times of a system consisting of several components, then $Q(y)$ can be viewed as the *rate of mortality on contour level y , given that all the components have not all failed within the n -coordinate time system up to level y* . Usually, the most efficient way of computing g , and hence Q , is to determine $V_{t,y}$ (and $dV_{t,y}/dy$) directly (rather than in surface integral form) and then use the relation

$$g(y) = \phi(y) \cdot \frac{dV_{t,y}}{dy}.$$

4. EXAMPLES AND COMPARISONS

4.1 Exponential-Type Multivariate Distributions.

a) De Simoni's (1968) family of multivariate distributions is given by

$$h(\underline{x}) = c \cdot e^{-[(\underline{x}-\underline{\mu})^T \Lambda (\underline{x}-\underline{\mu})]^a}$$

where $c > 0$, Λ is a positive definite matrix, $\mu \in \mathbb{R}^n$, $a > 0$. Accordingly, $t(\underline{x}) = -\log(h(\underline{x})/c) = [(\underline{x}-\mu)^T \Lambda (\underline{x}-\mu)]^a$ and the volume is

$$\begin{aligned} V_{t,y} &= \text{vol}\{\underline{x} | t(\underline{x}) \leq y\} \\ &= \text{vol}\{\underline{x} | (\underline{x}-\mu)^T \Lambda (\underline{x}-\mu) \leq y^{1/a}\} \\ &= \frac{1}{\sqrt{\det \Lambda}} \cdot \frac{\pi^{n/2}}{\Gamma(1+n/2)} \cdot y^{n/2a}. \end{aligned}$$

Thus $\frac{dV_{t,y}}{dy} = \frac{n}{2a} \cdot \frac{1}{\sqrt{\det \Lambda}} \cdot \frac{\pi^{n/2}}{\Gamma(1+n/2)} \cdot y^{n/2a-1}$, and $\phi(y) = ce^{-y}$.

Consequently,

$$\begin{aligned} g(y) &= \phi(y) \frac{dV_{t,y}}{dy} \\ &= c' \cdot y^{(n/2a)-1} e^{-y} \quad (c' = [\Gamma(n/2a)]^{-1}) \\ G(y) &= c' \cdot \int_0^y s^{(n/2a)-1} e^{-s} ds. \end{aligned}$$

Using the notation $P(n/2a, y) = \int_y^\infty s^{(n/2a)-1} e^{-s} ds$ for a form of the incomplete gamma function, we obtain

$$Q(y) = \frac{g(y)}{1-G(y)} = \frac{y^{(n/2a)-1} e^{-y}}{P(n/2a, y)}.$$

Thus Q is also the ordinary hazard rate of a *univariate gamma distribution*.

b) The θ -Generalized Multivariate Normal family (Goodman and Kotz, 1973) has density

$$h(\underline{x}) = c \cdot e^{-\left(\|A(\underline{x}-\mu)\|\right)_\theta^\theta},$$

where $c > 0$, A is non-similar n by n matrix, $\mu \in \mathbb{R}^n$, $\theta > 0$, and

$$\|\underline{y}\|_\theta = \left(\sum_{i=1}^n |y_i|^\theta \right)^{1/\theta}.$$

Accordingly, $t(\underline{x}) = -\log(h(\underline{x})/c) = (||A(\underline{x}-\underline{\mu})||_{\theta})^{\theta}$ and

$$\begin{aligned} V_{t,y} &= \text{vol}\{\underline{x} | t(\underline{x}) \leq y\} \\ &= \text{vol}\{\underline{x} | ||A \cdot (\underline{x}-\underline{\mu})||_{\theta} \leq y^{1/\theta}\} \\ &= \frac{1}{\det A} \cdot \frac{[2\Gamma(1+e^{-1})]^n}{\Gamma(1+n\theta^{-1})} \cdot y^{n/\theta} \end{aligned}$$

with the derivative

$$\frac{dV_{t,y}}{dy} = \frac{n\theta^{-1}}{\det A} \cdot \frac{(2\Gamma(1+\theta^{-1})^n}{(1+n\theta^{-1})} \cdot y^{n/\theta-1}$$

$$\phi(y) = c\theta^{-y}.$$

Consequently, $g(y) = \phi(y) \frac{dV_{t,y}}{dy} = c' \cdot y^{(n/\theta)-1} \cdot e^{-y}$ where

$c' = 1/\Gamma(n\theta^{-1})$. Analogously to the previous case,

$$Q(y) = \frac{g(y)}{1-G(y)} = \frac{y^{(n/\theta)-1} e^{-y}}{P(n\theta^{-1}, y)}.$$

Again, we note that the contour hazard rate is the same as the ordinary univariate hazard rate of a gamma distribution (but with a parameter different from that in (a)).

c) The univariate standardized exponential density $h(x) = e^{-x}$, $x \geq 0$, yields $t(x) = x$ with $\frac{dV_{t,x}}{dx} = 1$. The contour hazard rate coincides in this case with the ordinary hazard rate, both equal to the constant 1. (See also Section 5.)

4.2 Power-Type Multivariate Distributions on an Infinite Range.

a) The Pareto type I multivariate family has density

$$h(\underline{x}) = c \left(\frac{x_1}{a_1} + \dots + \frac{x_n}{a_n} + \alpha \right)^{-b}$$

where $x_1 > a_1 > 0$; $c, b > 0$. Accordingly,

$$t(\underline{x}) = \left(\frac{h(\underline{x})}{c} \right)^b - \alpha = \sum_{i=1}^n \frac{x_i}{a_i},$$

and the volume is $v_{t,y} = a_1 \cdots a_n \cdot \frac{(y-n)^n}{n!}$. Thus

$$\frac{dv_{t,y}}{dy} = a_1 \cdots a_n \cdot \frac{(y-n)^{n-1}}{(n-1)!} \quad \text{and} \quad \phi(y) = c \cdot (y+\alpha)^{-b}.$$

Consequently,

$$\begin{aligned} g(y) &= \phi(y) \cdot \frac{dv_{t,y}}{dy} \\ &= c' \cdot (y+\alpha)^{-b} \cdot (y-n)^{n-1}, \quad y \geq n. \end{aligned}$$

The constant c' is determined from the fact that g is a univariate density function. This leads to a computable integral for $1-G(y)$; and we note that g is a *translated form of the univariate F-density*. Hence Q will coincide with the ordinary hazard rate of this variable. In the particular case when the parameter α is of the "natural" form $-n+1$ and $b = a+n$, we have

$$g(y) = [B(a,n)]^{-1} (y-n+1)^{-(a+n)} (y-n)^{n-1}, \quad y \geq n,$$

and (by repeated integration by parts)

$$1-G(y) = \sum_{j=0}^{n-1} \frac{(n-1)^{[j]}}{(a+n-1)^{[j+1]}} \frac{(y-n)^{n-j-1}}{(y-n+1)^{a+n-j-1}},$$

where $x^{[j]} = x(x-1)\cdots(x-j+1)$; $x^{[0]} = 1$. Thus, in this case, Q is the "ordinary" hazard rate of a *univariate F-density* with parameters $2n$ and $2a$ which has been shifted to the right by a value n .

b) The multivariate t distribution has density

$$h(\underline{x}) = c \cdot ((\underline{x}-\underline{\mu})^T \Lambda (\underline{x}-\underline{\mu}) + a)^{-b}$$

where $c, b > 0$, a is real, Λ is a positive definite n by n matrix, and $\underline{\mu} \in R^n$. Accordingly,

$$t(\underline{x}) = \left(\frac{h(\underline{x})}{c} \right)^b - a$$

$$\epsilon = (\underline{x} - \mu)^T \Lambda (\underline{x} - \mu),$$

and the volume is $v_{t,y} = \frac{1}{\sqrt{\det \Lambda}} \cdot \frac{\pi^{n/2}}{\Gamma(1+n/2)} y^{n/2}$. Hence

$$\frac{dv_{t,y}}{dy} = \frac{1}{\sqrt{\det \Lambda}} \cdot \frac{n\pi^{n/2}}{2\Gamma(1+n/2)} y^{(n/2)-1},$$

and $\phi(y) = c(y+a)^{-b}$

Thus,
$$g(y) = \phi(y) \cdot \frac{dv_{t,y}}{dy}$$

$$= c' \cdot (y+a)^{-b} \cdot y^{(n/2)-1}.$$

The contour hazard rate is similar to that in a) but the parameters are different (due to the presence of the quadratic form rather than the linear one in the expression for the density).

4.3 Mixed Type (Power-Exponential) Distributions Defined on an Infinite Domain.

a) The generalized univariate gamma distribution has density

$$h(x) = c x^\alpha e^{-\beta x^\gamma}, \quad x \geq 0,$$

where $c, \beta, \gamma > 0; \alpha > -1$. Accordingly,

$$t(x) = -\log\left(\frac{h(x)}{c}\right) = \beta x^\gamma - \alpha \log x.$$

Then:
$$v_{t,y} = \text{vol}\{x | x_0 \leq x \leq t_1^{-1}(y)\}$$

$$+ \text{vol}\{x | t_2^{-1}(y) \leq x \leq x_0\}$$

$$= t_1^{-1}(y) - t_2^{-1}(y),$$

where x_0 satisfies the relation: $\log x_0 = \frac{\beta^\gamma}{\alpha} x_0^{\gamma-1}$ and t_1^{-1} is the increasing branch of t^{-1} over $[x_0, +\infty)$ while t_2^{-1} is a decreasing branch over $[x_0, +\infty)$. It follows easily that

$$\begin{aligned} \frac{dv_{t,y}}{dy} &= \frac{dt_1^{-1}(y)}{dy} - \frac{dt_2^{-1}(y)}{dy} \\ &= \frac{1}{\left(\frac{dt(x)}{dx}\right)_{x=t_1^{-1}(y)}} - \frac{1}{\left(\frac{dt(x)}{dx}\right)_{x=t_2^{-1}(y)}} \end{aligned}$$

while $\phi(y) = c e^{-y}$. Thus

$$g(y) = \phi(y) \frac{dv_{t,y}}{dy} = c' e^{-y} \cdot \frac{dv_{t,y}}{dy}.$$

Consequently,

$$\begin{aligned} 1-G(y) &= \int_{s=y}^{\infty} g(s) ds = c' \int_{s=y}^{\infty} e^{-s} \cdot \frac{dv_{t,s}}{ds} ds \\ &= c' \int_{s=y}^{\infty} e^{-s} \cdot \left(\frac{dt_1^{-1}(s)}{ds} - \frac{dt_2^{-1}(s)}{ds} \right) ds \\ &= c' \left(\int_{s=y}^{\infty} e^{-s} \cdot \frac{1}{\left(\frac{dt(u)}{du}\right)_{u=t_1^{-1}(s)}} ds \right. \\ &\quad \left. - \int_{s=y}^{\infty} e^{-s} \cdot \frac{1}{\left(\frac{dt(u)}{du}\right)_{u=t_2^{-1}(s)}} ds \right). \end{aligned}$$

The two integrals on the preceding line are found to be

$$\int_{x=t_1^{-1}(y)}^{\infty} x^{\alpha} e^{-\beta x^{\gamma}} dx \quad \text{and} \quad \int_{x=0}^{t_2^{-1}(y)} x^{\alpha} e^{-\beta x^{\gamma}} dx,$$

respectively. Denoting $\Gamma(\alpha, x) = \int_x^{\infty} e^{-t} t^{\alpha-1} dt$ with $\Gamma(\alpha, 0) = \Gamma(\alpha)$, we obtain after straightforward calculations

$$Q(y) = g(y)/[1-G(y)] = e^{-y} \cdot A \cdot (B_1^{-1} - B_2^{-1}) / (C_1 + C_2 - D)$$

where $A = \gamma \cdot \beta^{(\alpha+1)/\gamma}$,

$$B_i = \beta \gamma [t_i^{-1}(y)]^{\gamma-1} - \alpha / t_i^{-1}(y), \quad i=1,2,$$

$$C_i = \Gamma([\alpha+1]/\gamma, \beta \cdot [t_i^{-1}(y)]^{\gamma}), \quad i=1,2,$$

$$D = \Gamma([\alpha+1]/\gamma).$$

In the particular case of a gamma density given by $h(x) = x e^{-x}$, $x \geq 0$, we have $t(x) = x - \log(x)$ and $\phi(y) = e^{-y}$. Let $x_i = t_i^{-1}(y)$, $i=1,2$, be the inverse functions of $y = t(x)$ on either side of $x = 1$. Thus,

$$dV_{t,y}/dy = x_1/(x_1-1) - x_2/(x_2-1)$$

and hence the contour hazard rate is explicitly given by the formula

$$Q(y) = \frac{[x_1/(x_1-1) - x_2/(x_2-1)] e^{-y}}{(x_1-1) e^{-x_1} + (x_2-1) e^{-x_2} - 1}.$$

In this case of a non-monotonic univariate distribution, it is still meaningful to talk about the life failure between a and b , $0 < a < b < \infty$, corresponding to the inspection of the random component when we don't have information about an earlier or a late failure, yet we know that no failure has occurred between the times a and b . The contour set collapses here into a pair of points since a typical contour set is

$$\{t_1^{-1}(y), t_2^{-1}(y)\}.$$

5. CHARACTERIZING PROPERTIES OF CONSTANT AND INCREASING CONTOUR HAZARD RATES

The following question is of interest. Under a logarithmic (or a power) transform for obtaining t , what is the family of densities h whose contour hazard rates coincide with the ordinary hazard rate of a given univariate density f ? In particular, what is the family which corresponds to the "pivotal" density $f(y) = e^{-y}$, $y \geq 0$, whose ordinary hazard rate is the constant 1? The problem is of practical interest since it yields a systematic procedure of generating families of multivariate distributions with desired properties.

Consider the equation $g(y) = f(y) \equiv e^{-y}$ or, equivalently,

$$\phi(y) \cdot \frac{dV_{t,y}}{dy} = e^{-y}.$$

$$\text{Thus } \frac{dV_{t,y}}{dy} = \frac{e^{-y}}{\phi(y)}. \quad (3)$$

Once the t 's satisfying the above equation are determined, the h 's are immediately obtained as $h(\underline{x}) = \phi(t(\underline{x}))$.

a) For the logarithmic transforms:

$$h(\underline{x}) = \phi(g(\underline{x})) = c e^{-t(\underline{x})},$$

$$\text{and } t(\underline{x}) = -\log[h(\underline{x})/c], \quad \phi(y) = c \cdot e^{-y}.$$

Thus (3) becomes

$$\frac{dV_{t,y}}{dy} = \frac{e^{-y}}{c \cdot e^{-y}} = \frac{1}{c} \quad (4)$$

and hence

$$V_{t,y} = (1/c)y + d. \quad (4')$$

b) Analogously for a power transform we have:

$$h(\underline{x}) = \phi(t(\underline{x})) = c(t(\underline{x})+a)^{-(\alpha-1)}$$

and
$$t(\underline{x}) = \left(\frac{h(\underline{x})}{c} \right)^{(\alpha-1)} - a.$$

Thus (3) becomes

$$\frac{dV_{t,y}}{dy} = \frac{e^{-y}}{\phi(y)}, \quad (5)$$

and hence,

$$\frac{dV_{t,y}}{dy} = \frac{1}{c} (y+a)^{(\alpha-1)} e^{-y}. \quad (5')$$

If h is a *one-dimensional* density and t is differentiable, non-negative and vanishes at some finite point and has isolated maxima or minima, if any, (with $V_{t,y}$ always finite and satisfying the conditions stipulated in Theorem 1), then it easy to see that $dV_{t,y}/dy$ is the sum of "length" terms of the form

$$\frac{dt_j^{-1}(y)}{dy} - \frac{dt_{j-1}^{-1}(y)}{dy}$$

where t_j^{-1} is an increasing inverse branch of t to the right of the local minima and t_{j-1}^{-1} is the decreasing inverse branch to the left; (at an end point these terms may possibly be replaced by a suitable positive constant). Equivalently, a typical term in the sum is

$$\frac{1}{\left| \left(\frac{dt(s)}{ds} \right)_{s=t_j^{-1}(y)} \right|} + \frac{1}{\left| \left(\frac{dt(s)}{ds} \right)_{s=t_{j-1}^{-1}(y)} \right|},$$

where the summation is over *all* the minima. Clearly, in a neighborhood of an extremum these sums grow indefinitely while in equation (4) the constant $1/c$ is bounded for all y . This contradiction yields that, under the assumptions above, t must be monotonically increasing. This implies that $\frac{dt^{-1}(y)}{dy} = 1/c$

and thus, $t(y) = cy$. (The boundary condition assures that $d = 0$ in (4').) Hence we have shown that the family of univariate densities given by $h(x) = c \cdot e^{-cx}$, $c > 0$, $x \geq 0$ is the *only* family of one-dimensional densities with a constant contour hazard rate (under logarithmic transform). This is also a characterizing property of the ordinary hazard rate.

In the case of two or more dimensions, an analogous reasoning shows that t satisfying (4) cannot have isolated local minima (the rate of change of the volume under the curve at these points is unbounded). Therefore a class of functions t satisfying (4) in the case of n -dimensional densities under a logarithmic transform is given by $t(\underline{x}) = (\underline{\gamma}^T \underline{x})^n$ where the positive vector $\underline{\gamma}$ satisfies

$$|\underline{\gamma}| = \prod_{i=1}^n \gamma_i = c/n!$$

but is otherwise arbitrary. Indeed we have

$$\begin{aligned} V_{t,y} &= \text{vol}\{\underline{x} | (\underline{\gamma}^T \underline{x})^n \leq y\} \\ &= \text{vol}\{\underline{x} | \underline{\gamma}^T \underline{x} \leq y^{1/n}\} = \frac{1}{n! |\underline{\gamma}|} y = c^{-1} y. \end{aligned}$$

Thus $h(\underline{x}) = c \cdot e^{-(\underline{\gamma}^T \underline{x})^n}$ is a class of n -dimensional densities having a constant contour hazard rate (under the logarithmic transform for t). Observe that for $n \geq 2$ the multivariate

exponential density $h(\underline{x}) = c \cdot e^{-\underline{\gamma}^T \underline{x}}$ for any $\underline{\gamma} \geq 0$ does not generally possess a constant contour hazard rate. (Compare with example (b) in Section 4.1.) This is in some variance with the conclusion reached for the case of a *vector-valued* multivariate hazard rate as defined by Johnson and Kotz (1975b). Analogously, one can determine the structure of one-dimensional or multi-dimensional densities h with an ICHR (increasing contour hazard rate) and DCHR (decreasing contour hazard rate). We shall briefly discuss the ICHR case under the logarithmic transform.

We have $h(\underline{x}) = c \cdot e^{-t(\underline{x})}$ and, as before,

$$g(y) = c \cdot e^{-y} \frac{dV_{t,y}}{dy}.$$

Hence $Q(y) = \frac{g(y)}{1-G(y)} = \frac{e^{-y} \cdot \frac{dV_{t,y}}{dy}}{\int_{s=y}^{\infty} e^{-s} \cdot \frac{dV_{t,s}}{ds} ds}$.

Now, $\frac{dQ(y)}{dy} \geq 0$ if and only if

$$\int_{s=y}^{\infty} e^{-s} \left(\frac{dV_{t,s}}{ds} \right) ds \cdot \left(\frac{d^2 V_{t,y}}{dy^2} - \frac{dV_{t,y}}{dy} \right) + e^{-y} \cdot \left(\frac{dV_{t,y}}{dy} \right)^2 > 0.$$

A sufficient condition for this is that $\frac{d^2 V_{t,y}}{dy^2} > \frac{dV_{t,y}}{dy}$. The

latter is satisfied by any real valued differentiable function r on R^+ such that

$$r(y) + \frac{dr(y)}{dy} > 0 \quad \text{and} \quad \frac{dV_{t,y}}{dy} = r(y)e^y. \quad (6)$$

However the class of t 's satisfying (6) is, by the above argument, represented as

$$t(\underline{x}) = w^{-1}((\underline{c}^T \underline{x})^n), \quad \text{where } \underline{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} > 0,$$

$$|\underline{c}| = 1/n! \quad \text{and} \quad w(y) = \int_{s=0}^y r(s) \cdot e^s ds.$$

(w is a monotonically increasing function.) Consequently, a family of densities h with an ICHR is given by

$$h(\underline{x}) = c \cdot e^{-w^{-1}((\underline{c}^T \underline{x})^n)}.$$

We now state briefly the corresponding results for the case of densities under *power-type transforms*. Since the function in the r.h.s. of (5) is uniformly bounded, the above reasoning concerning the branches of the "inverse" of t is valid in this case as well and we have as above that, under mild restrictions on t , the *only* one-dimensional t 's satisfying (5') are monotonically increasing given by

$$\frac{dt^{-1}(y)}{dy} \equiv \frac{1}{c} (y+a)^{\alpha-1} e^{-y}, \quad y \geq -a.$$

Thus $t^{-1}(y) = \frac{e^a}{c} \cdot [\Gamma(a, \alpha) - \Gamma(a+y, \alpha)] = \psi(y)$, say. Since $\psi(\cdot)$ is a monotone increasing function, $t(y) = \psi^{-1}(y)$ and $h(x) = c\{\psi^{-1}(x)+a\}^{-(\alpha-1)}$, $x \geq 0$, is the *only* one-dimensional family of densities having a constant contour hazard rate under the power-type transformation.

An analogous argument for the n -dimensional case yields $h(\underline{x}) = c \cdot \psi^{-1}((\underline{c}^T \underline{x})^n + a)^{-(\alpha-1)}$ (where the vector $\underline{c} > 0$ satisfies $|\underline{c}| = \prod_{i=1}^n c_i = 1/n!$ but is otherwise arbitrary) as a family of densities with a constant contour hazard rate under the power-type transform. *More generally*, let the relation between h and t be given by $h(\underline{x}) = \phi(g(\underline{x}))$, where ϕ is a suitably chosen positive function (from R^1 to R^+). Setting

$$g(y) = \phi(y) \frac{dv_{t,y}}{dy} \equiv e^{-y},$$

we obtain

$$v_{t,y} = \int_0^y [\phi(s)]^{-1} e^{-s} ds = \chi(y), \text{ say,}$$

an increasing function of y . Choosing $t(\underline{x}) = \chi^{-1}((\underline{c}^T \underline{x})^n)$, where the vector \underline{c} is as above, we have

$$h(\underline{x}) = \phi(\chi^{-1}((\underline{c}^T \underline{x})^n))$$

as a family of n -dimensional densities with a constant contour hazard rate under the transform ϕ . Finally, note that in the case of $n \geq 2$ this family can be extended by choosing in place

of the form $(c^T \underline{x})^n$ a more general function of \underline{x} , $k(\underline{x})$, say, satisfying

$$\int \dots \int_{\{\underline{x} | k(\underline{x}) \leq s\}} d\underline{x} = s \quad \text{for all } s \geq 0.$$

For example, $k(\underline{x}) \equiv V_{m, m(\underline{x})}$ for a suitable $m: R^n \rightarrow R^+$, where as above, $V_{m, m(\underline{x})} = \text{vol}\{\underline{u} | \underline{u} \in R^n \text{ and } m(\underline{u}) \leq m(\underline{x})\}$, will be an appropriate choice.

6. SUMMARY

While the "classical" hazard rate is concerned with the behavior of a r.v. over the points of R^+ directly, the *contour* definition is determined through *sets of points* (i.e., the contours).

All one-dimensional concepts for hazard rates such as IHR, DHR, IHRA, and DHRA carry over immediately to multivariate densities by means of a corresponding univariate density g and comparisons of contour hazard rates - such as for IHR or DHR - can be accomplished for multivariate densities with a common contour set. Characterizing properties discussed in Section 5 indicate that the contour and ordinary hazard rates are compatible as measures of IHR and DHR properties and thus justify the choices of the "volume reducing" transformations suggested in this paper.

The proposed concepts may thus serve as a tool in studying various "hidden" properties and dependence interrelations of multivariate distributions and should be useful for generation of new families possessing desirable modelling features.

ACKNOWLEDGEMENTS

Dr. S. Kotz's research was supported by the U. S. Air Force Office of Scientific Research, Grant No. 78-3673.

REFERENCES

- Basu, A. P. (1971). Bivariate failure rate. *Journal of the American Statistical Association*, 66, 103-104.

- Bildikar, S. and Patil, G. P. (1968). Multivariate exponential-type distributions. *Annals of Mathematical Statistics*, 39, 1316-1326.
- Block, H. W. (1973). Monotone hazard and failure rates for absolutely continuous multivariate distributions. Department of Mathematics Research Report #73-20, University of Pittsburgh. Also *Naval Research Logistics Quarterly*, 24, 1977, 627-638.
- Brindley, E. C. and Thompson, W. A. (1972). Dependence and aging aspects of multivariate survival. *Journal of the American Statistical Association*, 67, 822-830.
- Day, N. E. (1969). Linear and quadratic discrimination in pattern recognition. *IEEE Transactions on Information Theory*, 15, 419-421.
- Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47, 307-323.
- Fraser, D. A. S. (1963). *The Structure of Inference*. Wiley, New York.
- Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Ser. 3*, 14, 53-77.
- Freund, R. J. (1961). A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56, 971-977.
- Fu, J. C. (1974). On non-uniqueness of ancillary statistics. *Communications in Statistics*, 3, 1093-1100.
- Goodman, I. and Kotz, S. (1973). Multivariate θ -generalized normal distributions. *Journal of Multivariate Analysis*, 3, 204-219.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698-707.
- Harris, R. (1970). A multivariate definition for increasing hazard rate functions. *Annals of Mathematical Statistics*, 41, 713-717.
- Higgins, J. J. (1975). A geometrical method of constructing multivariate densities and some related inferential procedures. *Communications in Statistics*, 4, 955-966.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics, Continuous Multivariate Distributions*. Wiley, New York.
- Johnson, N. L. and Kotz, S. (1975a). On some generalized Farlie-Gumbel-Morgenstern distributions. *Communications in Statistics*, 4, 415-428.
- Johnson, N. L. and Kotz, S. (1975b). A vector multivariate hazard rate. *Journal of Multivariate Analysis*, 5, 53-66.
- Marshall, A. W. (1975). Some comments on the hazard gradient. *Stochastic Processes and their Applications*, 3, 293-300.

- Marshall, A. W. and Olkin, I. (1967a). A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4, 291-302.
- Marshall, A. W. and Olkin, I. (1967b). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.
- Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsblatt für Mathematische Statistik*, 8, 234-235.
- Sagrasta, S. N. (1952). On a generalization of Pearson's curves to the two dimensional space. *Trabajos de Estadística*, 3, 273-314.
- Simoni, de S. (1968). Su una estensione delle curve normali di ordina r alle variabili doppie. *Statistica* (Bologna), 28, 151-170.
- Styen, H. S. (1960). On regression properties of multivariate probability functions of Pearson's type. *Proceedings of the Royal Academy of Sciences, Amsterdam*, 63, 302-311.
- Uven, M. J. van. (1947-1948). Extension of Pearson's probability distributions to two variables, I-IV. *Proceedings of the Royal Academy of Sciences, Amsterdam*, 50, 1063-1070, 1252-1264; 51, 41-52, 191-196.

[Received March 1980. Revised September 1980]

FAILURE TIME DISTRIBUTIONS: ESTIMATES AND ASYMPTOTIC RESULTS

JANOS GALAMBOS

Department of Mathematics
Temple University
Philadelphia, Pennsylvania 19122 USA

SUMMARY. The paper deals with life distributions for coherent systems of components. Two major questions are discussed: (i) estimation of system life in terms of component lives, and (ii) asymptotic models. Both questions are related to extremes of a sequence of random variables through the path set and cut set decomposition of coherent systems, which reduce a coherent system to either a parallel or a series system. It is pointed out that the classical theory of extremes of independent and identically distributed random variables does not provide an acceptable approximation. Hence, the emphasis is on dependence or on the case when the random variables are not identically distributed. The inequalities presented when discussing question (i) are applicable not only to extreme value problems but to an arbitrary multivariate distribution when lower dimensional marginals are specified. The asymptotic models are also discussed in the light of hazard rate properties of the limiting distributions for the extremes in a particular model.

KEY WORDS. coherent system, component life, system life, extreme value distribution, dependent model, inequalities, multivariate distribution, Weibull distribution, hazard rate.

1. LIFE DISTRIBUTION OF COHERENT SYSTEMS: THE REDUCTION TO EXTREME VALUE DISTRIBUTIONS

In this introductory section we collect some definitions which are needed to present two general representations of coherent

systems, the so called minimal path set and minimal cut set representations. These two representations lead to the fact that the life distribution (failure time distribution) of every coherent system is an extreme value distribution in some dependent model of random variables. The dependence is the major emphasis here; the classical extreme value model with independent and identically distributed random variables is very rarely applicable in reliability theory.

Let a system consist of n components. We denote by X_j and T the random life length of the j th component and the system, respectively. Evidently, the events $\{X_j > x\}$ and $\{T > x\}$ express the fact that the j th component and the system function at time x . The following assumptions are made throughout this paper: (i) the event $\{T > x\}$ is uniquely determined by the events $\{X_j > x\}$, $1 \leq j \leq n$, or by their complements; (ii) if $\{T > x\}$ in the case $\{X_j \leq x\}$ for some j , then T remains larger than x if the j th component is replaced by a functioning similar component; and (iii) $\{T > x\}$ does depend on $\{X_j > x\}$ for each j . A system with these three properties is called a coherent system with each of its components essential. See Chapters 1 and 2 of Barlow and Proschan (1975) for the basic properties of coherent systems.

There are two special systems which will play specific roles in the sequel and which we define below.

Series system. If a system fails as soon as one of its components fails, we call it a series system. Hence,

$$T = \min(X_1, X_2, \dots, X_n).$$

Parallel system. We call a system parallel, if it functions as long as one of its components functions. It follows that

$$T = \max(X_1, X_2, \dots, X_n).$$

We shall now show that an arbitrary coherent system can be reduced to a parallel or a series system by a suitable grouping of the components. For proving this fact, we introduce the following concepts.

Minimal path sets. A path set of a coherent system is a set C of components such that if each member of C functions then the system functions. A path set is minimal, if the removal of a single element from it results in its failure to remain a path set.

Minimal cut sets. A set C of components is called a cut set if the system fails whenever each member of C fails. A cut set is minimal if no element can be removed from it without violating its cut set property.

Now a coherent system which is capable of functioning is necessarily both a path set and a cut set. Therefore, every such coherent system has at least one minimal path set and one minimal cut set. Let A_1, A_2, \dots, A_p and C_1, C_2, \dots, C_m be the distinct minimal path and cut sets, respectively, for a given coherent system. Then, by definition

$$T = \max_{1 \leq t \leq p} \left\{ \min_{j \in A_t} X_j \right\} = \min_{1 \leq t \leq m} \left\{ \max_{j \in C_t} X_j \right\}.$$

$$\text{Putting } U_t = \min_{j \in A_t} \{X_j\}, \quad V_t = \max_{j \in C_t} \{X_j\}, \quad (1)$$

$$\text{we obtain } T = \max_{1 \leq t \leq p} \{U_t\} = \min_{1 \leq t \leq m} \{V_t\}.$$

In other words, the system life of an arbitrary coherent system is an extreme value of a suitably chosen sequence of random variables. Its distribution function, which we call a failure time distribution (for coherent systems), is therefore an extreme value distribution in some appropriate model. It should be emphasized that the independence and stationarity cannot be assumed here even if the original components are believed to function independently (such a case is rare though; see the last but one paragraph of the next section in this regard). Because the minimal path sets A_t (or the minimal cut sets C_t) are not disjoint, the random variables U_t (or V_t) of (1) are strongly dependent.

Their dependence is determined by the structure of the underlying system and by the dependence of the original components. Hence, their dependence is never a matter of arbitrary assumptions. Thus the study of these distributions is an integral part of the theory of the extremes for dependent models. While such a theory is well developed in Chapter 3 of the present author's book, Galambos (1978), we shall discuss some points of this theory as they relate to failure time distributions. Although in the previous paragraph we defined failure time distributions in terms of coherent systems, much of what is to be said also applies to failure times of dams, when failure is caused by high floods; of air quality, when failure is defined by the fact that some specific pollutant exceeds a given level of concentration, and to other diverse fields. We shall, however, remain in our discussion

at coherent systems, and we use the notations and concepts of the present section throughout the paper.

2. ASYMPTOTIC EXTREME VALUE DISTRIBUTIONS AS FAILURE TIME DISTRIBUTIONS

We have seen that T can always be expressed as the maximum or the minimum of some random variables. Since in these representations either p or m is large for a system with a large number n of components (recall that we consider essential components only), we assume for the general discussion that

$$T = \max_{1 \leq t \leq p} \{U_t\}$$

with p large. Now if the distributions and the interdependence of the U_t are known then the distribution of T is uniquely determined and can be computed by routine methods. Hence, only that case presents a problem when either the distribution functions $F_t(x) = P(U_t < x)$, $1 \leq t \leq p$, or the dependence relation of the U_t are unknown. In such cases, an approximation becomes necessary. It is shown on p. 90 of Galambos (1978) (and further explored in Galambos, 1981) that a reliable approximation to the distribution of T cannot be obtained through an approximation to $F_t(x)$. Rather, one should develop a dependent model for the U_t , evaluate the possible limiting distributions for the maximum in that model, and one of these possibilities is to be applied as an approximation to the distribution of T .

There are a number of dependent models for which the mathematical results are at an advanced stage (although far from complete). These are described in Chapter 3 of the mentioned book of the present author, from which we quote the following results.

- (i) Approximation by the classical model (which would assume that the U_t are independent and identically distributed) is very rarely justified, but when it is applicable, then the failure time distribution is Weibull.
- (ii) The assumption that the U_t form a sequence of exchangeable random variables is mathematically justified for an arbitrary coherent system. But, because of this generality, the possible limit laws for the maximum form a very large family. The investigation of the properties of some subfamilies of these distributions would be a very important

task. It should be remarked that, contrary to the claim in the paper Zidek *et al.* (1979), the sequence U_t cannot, in general, be considered as a segment of an infinite sequence of exchangeable variables; only finite exchangeability is justified.

- (iii) There is a general dependent model, in which the possible limit laws for the maximum coincide with the set of those distributions whose hazard rate function (defined below) is monotonic (see Sections 3.9 and 3.10 in the quoted book). Because of the significance of this result to engineers, we discuss this conclusion in more detail.

Engineers have recognized for a long time the importance of the hazard rate function of a failure time distribution. Let us first give the definition of hazard rate. Let $X > 0$ be a random variable with distribution function $F(x)$ and with density function $f(x) = F'(x)$. Then the hazard rate $r(x)$ of X is defined by the limit relation

$$r(x) = \lim_{\delta x \rightarrow 0} \frac{1}{\delta x} P(X < x + \delta x | X \geq x).$$

An easy calculation yields from this limit relation

$$r(x) = f(x)/[1-F(x)]. \quad (2)$$

While this latter form is a convenient formula for actually calculating $r(x)$, the definition of $r(x)$ is what makes it applicable, since it represents the rate at which X fails in a short time interval $(x, x + \delta x)$, given that X has survived beyond x .

It is apparent to an engineer that a new system of components whose life is represented by X may have some positive probability of failing immediately after production, but this probability decreases as time passes (burn-in period). On the other hand, an old system of components is more and more likely to fail as time passes (aging or wear-out period). Between the burn-in and wear-out periods, it is accepted that, for most systems, only accidents may cause failure (accidental failure period). The stochastic definition of accidents is either by constant hazard rate or by the lack of memory property $P(X \geq x+y | X \geq x) = P(X \geq y)$. See Section 1.5 in Galambos and Kotz (1978) to see that these two seemingly different definitions are equivalent.

Since each of the three periods above represents a monotonic failure rate, it is very pleasing to see that the mathematical theory through structure functions and extreme value theory justifies the intuitive argument of the engineers.

Another important consequence of accepting the existence of an accidental failure period for a system is that it excludes the possibility for the components to function independently. Namely, if we write (2) in the form

$$r(x) = -d\{\ln(1 - F(x))\}/dx.$$

we then get by integration

$$F(x) = 1 - \exp\left\{-\int_0^x r(t) dt\right\}, x > 0.$$

We thus see that when the hazard rate is constant, then $F(x)$ is exponential. In particular, during the accidental failure period, life distributions are always exponential. Assume now that all components as well as the system achieved the accidental failure period. This means, that both the components and the system have exponential failure distributions (for a certain period of time only). It then follows from a result of Esary *et al.* (1971) that the components, with the exception of series systems, are stochastically dependent. That is, *one cannot construct a single structure other than a series system in which the components would function independently* (and which system would achieve an accidental failure period). This is a very important conclusion because several estimates on reliability are developed in the literature under the assumption that the components are independent.

Finally, we remark that there are a number of characterization theorems for exponentiality (see the book Galambos and Kotz, 1978) which can be used for testing whether a system is in its accidental failure period. In most cases, those limited characterization theorems are sufficient when one assumes a priori that the underlying distribution is of monotonic hazard rate. A typical result of this nature can be found in Ahsanullah (1977).

3. ESTIMATES ON FAILURE TIME DISTRIBUTIONS

We have emphasized in the previous section that components for most structures cannot function independently of each other. At the same time, we may know exactly the distribution of component lives, mainly through characterization theorems. This leads us to the problem of estimating failure time distributions by the distributions of component lives under some assumption of dependence of the components.

We use the path set and cut set decompositions, in view of which structure life is an extreme of "component lives" (where "component" is either a minimal path set or a minimal cut set).

Through this approach, the mathematical problem is the estimation of the distribution function $H_n(x)$ of

$$T = \max(U_1, U_2, \dots, U_p)$$

under some form of dependence of the U_t and under the assumption that the distribution functions $F_t(x) = P(U_t < x)$ are known.

There is one concept of dependence, the so-called association of random variables, for which there is an extensive literature with reliability emphasis (see Barlow and Proschan, 1975, and Natvig, 1980). However, these works deal with estimating $E(T)$ in terms of $E(U_t)$, $1 \leq t \leq p$, rather than giving estimates on $H_n(x)$.

Since we deal with distributions, we express dependence through distributional assumptions. The simplest distributional assumption is, of course, when only bivariate distributions are involved. For simplicity, we introduce the notations

$$A_j = A_j(x) = \{U_j \geq x\}$$

$$S_{1,p}(x) = S_{1,p} = \sum_{j=1}^p P(A_j),$$

$$S_{2,p}(x) = S_{2,p} = \sum_{1 \leq i < j \leq p} P(A_i \cap A_j),$$

and we let $m_p = m_p(x)$ stand for the number of those A_j which occur. Then $H_n(x) = P(m_p = 0)$, $S_{1,p} = E(m_p)$ and $2S_{2,p} + S_{1,p} = E(m_p^2)$. This latter meaning of $S_{1,p}$ and $S_{2,p}$ makes them appealing to the applied statistician, while their original definition is the useful form in mathematical arguments. It is slightly more convenient to state results for $1 - H_n(x) = P(m_p \geq 1)$ than for $H_n(x)$ itself, and we shall do so below. Let us consider estimates of the form

$$a S_{1,p} + b S_{2,p} \leq P(m_p \geq 1) \leq c S_{1,p} + d S_{2,p}, \quad (3)$$

where a, b, c and d are constants (which, in principle, may depend on x suppressed in all notations).

The best lower bound in (3) is known (Kwerel, 1975, and Galambos, 1977), according to which a and b should be of the form $a = 2/(k+1)$ and $b = -2/k(k+1)$, where $1 \leq k < p$ is an integer. It is then easy to find the optimal k which equals

$[2S_{2,p}/S_{1,p}] + 1$, where $[y]$ signifies the integer part of y . (Notice that $k = 1$ yields the classical estimate by the method of inclusion and exclusion.) For the upper bound in (3), only partial results are available. The best known result is $c = 1$ and $d = -2/p$ (Kounias, 1968, and Galambos, 1975).

Before proceeding with the discussion of the estimates in (3), notice the following important fact. The results quoted in the previous paragraph are such that the coefficients a , b , c and d do not depend on x . Hence, they remain valid if we redefine A_j .

Now, if $A_j = \{U_j \geq x_j\}$, then

$$\{m_p = 0\} = \{U_1 < x_1, U_2 < x_2, \dots, U_p < x_p\},$$

and thus (3) provides estimates on the p -variate distribution of the U_j in terms of univariate and bivariate marginals. These inequalities should be taken into account when one is interested in constructing multivariate distributions with given (univariate and bivariate) marginals.

Let us return to (3). Since $P(m_p \geq 1)$ is related to the distribution of the maximum of the U_j , one would like to get suitable extensions of (3) to $P(m_p \geq r)$, $r \geq 1$, which is relevant for the distribution of the $(p-r+1)$ st order statistic of the U_j . Two methods of proof of (3) lead to interesting results in this direction. One proof, which roughly says that (3) is valid for arbitrary (dependent) sequence of U_j if it is valid when the U_j are i.i.d., implies that a set of coefficients in (3) determines a set of coefficients for estimating $P(m_p \geq r)$ in the form of (3) (Galambos and Mucci, 1980). The other proof, introduced in Galambos (1977), provides a technique which can be used with success in more general cases than (3) (for example to $P(m_p \geq r)$ for any $r \geq 1$, and with bounds not necessarily linear). On this line of extensions of (3), we mention Sathe *et al.* (1980). For earlier results on inequalities of the nature of (3), see the survey at the end of Chapter 1 of Galambos (1978).

ACKNOWLEDGEMENTS

This research was financially supported by a grant from the Air Force Office of Scientific Research to Temple University (Grant Number AFOSR 78-3504 & 3504A). The author also wishes to acknowledge with appreciation the Research Leave granted by Temple University, during which this paper was completed.

REFERENCES

- Ahsanullah, M. (1977). A characteristic property of the exponential distribution. *Annals of Statistics*, 5, 580-582.
- Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston, New York.
- Esary, J. D., Marshall, A. W. and Proschan, F. (1971). Determining an approximate constant failure rate for a system whose components have constant failure rates. In *Operations Research and Reliability*, D. Grouchko, ed. Gordon and Breach, New York.
- Galambos, J. (1975). Methods for proving Bonferroni type inequalities. *Journal of the London Mathematical Society, Series 2*, 9, 561-564.
- Galambos, J. (1977). Bonferroni inequalities. *Annals of Probability*, 4, 577-581.
- Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York.
- Galambos, J. (1981). Extreme value theory in applied probability. *The Mathematical Scientist*, 6, 13-26.
- Galambos, J. and Kotz, S. (1978). *Characterizations of Probability Distributions*. Lecture Notes in Mathematics Series, Vol. 675, Springer Verlag, Heidelberg.
- Galambos, J. and Mucci, R. (1980). Inequalities for linear combinations of binomial moments. *Publicationes Mathematicae Debrecen*, 27, to appear.
- Kounias, E. G. (1968). Bounds for the probability of a union, with applications. *Annals of Mathematical Statistics*, 39, 2154-2158.
- Kwerel, S. M. (1975). Most stringent bounds on aggregated probabilities of partially specified dependent systems. *Journal of the American Statistical Association*, 70, 472-479.
- Natvig, B. (1980). Improved bounds for the availability and unavailability in a fixed time interval for systems of maintained, interdependent components. *Advances in Applied Probability*, 12, 200-221.
- Sathe, Y. S., Pradhan, M., and Shah, S. P. (1980). Inequalities for the probability of the occurrence of at least m out of n events. *Journal of Applied Probability*, to appear.
- Zidek, J. V., Navin, F. P. D. and Lockhart, R. (1979). Statistics of extremes; an alternate method with application to bridge design codes. *Technometrics*, 21, 185-191.

[Received June 1980. Revised October 1980]

A NOTE ON SHOCK MODEL JUSTIFICATION FOR IFR DISTRIBUTIONS

PURUSHOTTAM LAUD and ROY SAUNDERS

Department of Mathematical Sciences
Northern Illinois University
DeKalb, Illinois 60115 USA

SUMMARY. We consider the model in which the failure rate for a device changes when the device is subjected to shocks which occur stochastically over time. We show that increasing failure rate distributions can be obtained by making simple models for the effects of shocks. The results provide a physical motivation for using the Weibull distributions for failure time data. Random failure rates used in Bayesian inference are also obtained in a similar manner by modeling the effects of shocks to be stochastic.

KEY WORDS. Failure rate, Poisson process, shock model, stochastic failure rate, Weibull distribution, weak convergence.

1. INTRODUCTION

The problem of choosing a particular family of distributions for analyzing failure time data is difficult since in addition to requiring that some member of the family "fit" the data it is also desirable to have some physical justification for using the chosen family of distributions.

As evidenced in the historical section of Barlow and Proschan (1965) early work on justifications was primarily concerned with specific families of distributions such as the exponential and Weibull families. The typical approach in the early work consisted of deriving specific families by considering failure times of complex pieces of equipment containing numerous components and making various reasonable assumptions about the joint failure distributions of the components.

More recently Esary *et al.* (1973) have presented justifications for the use of more general families of distributions such as the increasing failure rate distributions. These justifications are based on a shock model which assumes that items are subject to shocks occurring as a Poisson process over time and that an item fails after receiving some random number of shocks. The justifications consist of showing that when the distributions of the number of shocks an item survives are of various forms then the failure rate must also have a certain general form such as being increasing. Generalizations of these justifications of Esary *et al.* to include different processes generating the shocks and random effects of shocks have been considered also.

In this note we use the basic idea from the shock model and apply it to failure rates in order to develop a novel method of justifying the use of some distributions widely used to analyze failure data. We also show how the method can be extended to provide justifications for some random failure rates processes which have been used by Dykstra and Laud (1980) in Bayesian analysis of failure data.

Specifically we assume that an item is subjected to shocks which, using $X(t)$ to denote the number of shocks occurring in $[0, t]$, occur as a Poisson process $\{X(t): t \geq 0\}$ with mean function $E\{X(t)\} = \lambda t$. Different from the Esary *et al.* model we assume that the shocks affect the failure rate of the item and that the failure rate of an item is a stochastic failure rate process $\{W(t): t \geq 0\}$ which is a function $W(t) = h_t(\{X(\tau): 0 \leq \tau \leq t\})$ of the shock process. By assuming two specific and intuitive forms for the functions $h_t(\cdot)$ we show that widely used failure distributions can be generated as limiting cases of this model when shocks occur frequently and each shock has a small effect on the failure rate. The results relating to Bayesian analysis are obtained in a similar manner when the effects of the individual shocks are allowed to be random variables as opposed to the fixed functions $h_t(\cdot)$.

2. RESULTS FOR AN ADDITIVE MODEL AND A MULTIPLICATIVE MODEL

The additive model assumes that the effect of successive shocks is to increase the failure rate for the item. That is, for some nonnegative real valued function $h(\cdot)$, the i th shock increases the failure rate by the amount $h(i)$ so that $W(t) = h(0) + \dots + h(X(t))$.

To consider the situation where numerous shocks occur in any time period and each shock has a small effect let $\{X_n(t): 0 \leq t < \infty\}$

denote a sequence of Poisson processes having mean functions $E\{X_n(t)\} = \lambda_n t$ and let $\{h_n : n=1,2,3,\dots\}$ be a sequence of shock effect functions for the additive model. Note that the number of shocks will be large if $\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$ and the effect of the shocks on the hazard rate will be individually small if $h_n(i) \rightarrow 0$ for all i as $n \rightarrow \infty$. With proper choices for these rates λ_n and functions $h_n(\cdot)$, the stochastic failure rates $W_n(t) = h_n(0) + \dots + h_n(X_n(t))$ converge to the deterministic failure rates of widely used survival distributions. For example if we take $\lambda_n = \delta n$, $h_n(0) = 0$, and $h_n(i) = \beta/n$ then, for each finite $\tau > 0$, the rates $\{W_n(t) : 0 \leq t \leq \tau\}$ converge in probability to the deterministic failure rate function $W_\infty(t) = \delta\beta t$. Thus the corresponding distribution of failure times converges to a particular member of the Weibull family with parameters $\alpha = 2$ and $\gamma = \delta\beta/2$ having density function

$$f(t) = \begin{cases} \gamma \alpha t^{\alpha-1} e^{-\gamma t^\alpha} & , \quad t \geq 0, \gamma > 0, \alpha > 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

and failure rate function $r(t) = \gamma \alpha t^{\alpha-1}$.

Other members of the Weibull family can be obtained similarly. to do this take $\lambda_n = \delta n$ with $\delta = \gamma^{1/(\alpha-1)}$, $h_n(0) = 0$ and $h_n(i) = \alpha[(i/n)^{\alpha-1} - ((i-1)/n)^{\alpha-1}]$ for $i \geq 1$. In this case we have $\sum_{0 \leq j \leq i} h_n(j) = \alpha(i/n)^{\alpha-1}$ and $W_n(t) = \alpha(X_n(t)/n)^{\alpha-1}$ which converges in probability to the deterministic failure rate $W_\infty(t) = \gamma \alpha t^{\alpha-1}$. It is interesting to note that when $\alpha = 2$ then $h_n(i+1) - h_n(i) = 0$ which implies constant shock effect while, for $\alpha = 3$, $h_n(i+1) - h_n(i) = 2h_n(1)$ which implies each shock affects the failure rate by two more units than the previous shock. In general $h_n(i+1) - h_n(i)$ is equal to $[(i+1)^{\alpha-1} + (i-1)^{\alpha-1} - 2i^{\alpha-1}]h_n(1)$ so that for $\alpha > 2$ the successive shocks have increasing effects while for $1 < \alpha < 2$ the successive shocks have decreasing effects.

A second way of modeling the shock effects is to assume that each shock increases the failure rate by a factor $1 + h(i)$ where $h(i)$ is again a nonnegative real valued function and the

stochastic failure rate process is given by $W(t) = h(0) \times (1 + h(1)) \times \dots \times (1 + h(X(t)))$. If for this multiplicative model we consider sequences as above then we also obtain well-known deterministic failure rates in the limit. The Weibull failure rate $\gamma \alpha t^{\alpha-1}$ is obtained by using $\lambda_n = \delta n$, $h_n(0) = n^{-\alpha}$, $h_n(i) = (1 + \alpha/i)$ for $i \geq 1$, and $\delta = [\gamma \alpha \Gamma(\alpha+1)]^{1/\alpha}$ since then $(\prod_{i=1}^{\infty} h_n(i) \equiv 1)$

$$W_n(t) = n^{-\alpha} \prod_{i=1}^{X_n(t)} \frac{\alpha + 1}{i} \approx (\delta X_n(t) + 1) (n\delta)^{\alpha-1} / n\alpha \Gamma(\alpha)$$

which converges in probability to $\delta^\alpha t^{\alpha-1} / \Gamma(\alpha+1) = \gamma \alpha t^{\alpha-1}$.

As a second example note that to obtain the limiting failure rate $r(t) = (\alpha/\gamma) \exp(\alpha t)$ for $t \geq 0$ which arises in the modified extreme value distribution (see, e.g., Barlow and Proschan, 1965) having density

$$f(t) = \begin{cases} \frac{\alpha}{\gamma} \exp(\alpha t - (e^{\alpha t} - 1)/\gamma) & , \quad t \geq 0, \quad \gamma > 0, \quad \alpha > 0 \\ 0 & , \quad \text{otherwise,} \end{cases}$$

choose $\lambda_n = \delta n$, $h_n(0) = \alpha/\gamma$, and $h_n(i) = \beta/n$ for $i \geq 1$, where $\delta\beta = \alpha$. Where it is possible to rewrite multiplicative models as additive models in most cases, one of the models should have more intuitive appeal in a given physical situation.

3. RESULTS FOR RANDOM SHOCK EFFECTS

In this section we consider an additive model in which each shock increases the failure rate by a random amount. Let $\alpha(t)$, $t \geq 0$, be a continuous increasing function of t with $\alpha(0) = 0$ and $\lim_{t \rightarrow \infty} \alpha(t) = M < \infty$ as $t \rightarrow \infty$. For each $n \geq 1$ let $H_n(i)$, $i \geq 1$, be a sequence of independent gamma random variables with shape parameters $\alpha(i/n) - \alpha((i-1)/n)$ and scale parameter 1 and let $H_n(0)$ be degenerate at 0. For each n let $\{X_n(t): t \geq 0\}$ be an homogeneous Poisson process with mean function nt . As in the preceding section we model the shocks to occur as events in $X_n(\cdot)$. The i th shock, however, increases the failure rate by a random amount $H_n(i)$. We take the shock effects random variables $\{H_n(i): i \geq 0\}$ to be independent of

the shock process $\{X_n(t): t \geq 0\}$. We thus obtain the stochastic failure rate $W_n(t) = H_n(0) + \dots + H_n(X_n(t))$. Now, as n increases, the number of shocks in any interval is large and each shock effect is stochastically small. In the limit the random variable $W_n(t)$ converges to a gamma random variable with scale parameter $\alpha(t)$ for each fixed t since the characteristic function of $W_n(t)$ is

$$\begin{aligned}\psi_n(\theta) &= E[\exp\{i\theta \sum_{j=1}^{X_n(t)} H_n(j)\}] \\ &= E\{E[\exp\{i\theta \sum_{j=1}^{X_n(t)} H_n(j)\} | X_n(t)]\} \\ &= E[\prod_{j=1}^{X_n(t)} \{1/(1-i\theta)\}^{\alpha(j/n)} \cdot \alpha((j-1)/n)] \\ &= E[\{1/(1-i\theta)\}^{\alpha(X_n(t)/n)}]\end{aligned}$$

which converges to $\{1/(1-i\theta)\}^{\alpha(t)}$. A multivariate extension of this argument establishes the convergence of the finite dimensional distributions of the sequence $\{W_n(\cdot)\}$ to those of a gamma process with parameter function $\alpha(\cdot)$. Using the definition and discussion of the space $D[0, \infty]$ in Rao and Sethuraman (1975) and a well-known sufficient condition (see Billingsley, 1968, Theorem 15.6) for weak convergence, it is straightforward to conclude weak convergence of $\{W_n(t): t \geq 0\}$ to a gamma process with parameter function $\alpha(\cdot)$. Reynolds and Savage (1971) employed gamma process and, more generally, an independent-increment stochastic process to describe random failure rates which are introduced by Gaver (1963) to explain failures of components subjected to random environments. Dykstra and Laud (1980) use the extended gamma process in Bayesian estimation of an increasing failure rate. The above random shock effects model can be used to obtain these random failure rates. Following Reynolds and Savage (1971), consider an independent increment stochastic process $\{Y(t): t \geq 0\}$ with no normal component, no deterministic component, and no fixed points of discontinuity. For such a process the logarithm of the characteristic function of $Y(t)$ has the Lévy representation (Notation: $E[\exp\{iuY(t)\}] = \exp\{\psi_t(u)\}$)

$$\psi_t(u) = \int_{0+}^{\infty} \{\exp(iuz) - 1\} d_z N(t, z)$$

where (i) N is a nondecreasing function of z and t with nonnegative second differences and continuous in t ; (ii) $N(t, \infty) = 0$ for $t \geq 0$; and

$$(iii)' \int_{0+}^{\infty} \frac{z}{1+z^2} d_z N(t, z) < \infty \text{ for } t \geq 0.$$

We also require that the mean function of the process is bounded. Thus (iii)' is replaced by

$$(iii) \int_{0+}^{\infty} z d_z N(t, z) \leq M < \infty \text{ for all } t \geq 0.$$

To obtain this random failure rate $\{Y(t): t \geq 0\}$, suppose that the random shock effect $H_n(j)$ has a distribution with characteristic function

$$E[\exp\{iu H_n(j)\}] = \exp\left[\int_{0+}^{\infty} \{\exp(iuz) - 1\} d_z \{N(\frac{j}{n}, z) - N(\frac{j-1}{n}, z)\}\right].$$

Again letting $W_n(t) = H_n(0) + \dots + H_n(X_n(t))$ where $H_n(0)$ is degenerate at zero, it can be shown, by using techniques similar to those used above to obtain the gamma random failure rate, that the sequence $\{W_n(t): t \geq 0\}$ converges weakly to $\{Y(t): t \geq 0\}$.

This provides a physical model justifying the random failure rates used by the above-mentioned authors.

4. CONCLUSION

The purpose of this article is to point out that when it is possible to envision shocks affecting the failure rate of an item, assumptions about the way the shocks affect the failure rate can lead to specific well-known models of the survival distributions or failure rates. We believe these results also can be used for approximating and suggesting priors for Bayesian nonparametric estimation of failure rate functions.

REFERENCES

- Barlow, R. E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Dykstra, R. L. and Laud, P. W. (1980). A Bayesian nonparametric approach toward reliability. *Annals of Statistics* (to appear).
- Esary, J. D., Marshall, A. W., and Proschan, F. (1973). Shock models and wear processes. *Annals of Probability*, 4, 627-649.
- Gaver, D. P., Jr. (1963). Random hazard in reliability problems. *Technometrics*, 5, 211-226.
- Rao, J. S. and Sethuraman, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbation and scale factors. *Annals of Statistics*, 3, 299-313.
- Reynolds, D. S. and Savage, I. R. (1971). Random wear models in reliability theory. *Advances in Applied Probability*, 3, 229-248.

[Received May 1980. Revised September 1980]

ON THE MEAN RESIDUAL LIFE FUNCTION IN SURVIVAL STUDIES

RAMESH C. GUPTA

Department of Mathematics
University of Maine at Orono
Orono, Maine 04469 USA

SUMMARY. In reliability studies, the expected additional life time given that a component has survived until time t is called the mean residual life function (MRLF). This MRLF determines the distribution function uniquely. In this paper an interpretation of MRLF in renewal theory is presented and some characterizations of the exponential distribution are obtained. Finally, considering the general MRLF, a method is developed for obtaining the mixing distribution when the original distribution is exponential. Some examples are discussed, in one of which Morrison's (1978) result is obtained as a special case.

KEY WORDS. mean residual life function, failure rate, survival analysis, characterizations, renewal process, mixture.

1. INTRODUCTION

In life testing situations, the expected additional life time given that a component has survived until time t is called the mean residual life function (MRLF). More specifically, if X is the life of a component, then $E(X - t | X > t)$ is called the MRLF. This MRLF has been employed in life length studies by various authors e.g. Hollander and Proschan (1975), Bryson and Siddiqui (1969) and Muth (1977). Limiting properties of the MRLF have been studied by Meilijson (1972) and Balkema and DeHann (1974). It has also been shown by the author (1975) that the MRLF determines the distribution function uniquely.

In reliability studies, the survival function is often characterized by its failure rate (hazard rate). An essential difference between the failure rate and the MRLF is that the former accounts only for the immediate future in assessing event component failure, whereas the latter accounts for the complete future. Thus in a sense MRLF should play as important a role as the failure rate.

Watson and Wells (1961) have found general conditions on a life distribution so that the mean remaining life of articles, operated for some fixed test period, is greater than the original mean life and, in this context, they have examined such well-known life distributions as the Weibull, gamma, lognormal and extreme value. Weiss and Dishan (1971) extend the work of Watson and Wells by assigning cost functions and by considering the economic aspects of burn-in and replacement. Muth (1977) focuses on the decreasing mean residual life functions and the properties of the associated probability distributions. He also defines the properties of negative memory, no memory, positive memory and perfect memory as associated with the MRLF, and examines several well-known distributions in terms of their memory properties.

After presenting some preliminary results in Section 2, we compare MRLF and failure rate in Section 3. It is shown that if the life distribution has increasing failure rate (IFR), then MRLF is decreasing but not conversely. Some comparisons of total life and MRLF are also mentioned. In Section 4, an interpretation of MRLF in renewal theory is presented and some characterizations of the exponential distributions are obtained. More specifically, we show that (i) if the renewal distribution has IFR and the mean life is equal to the mean residual life, then the distribution is exponential and (ii) if the renewal distribution belongs to the one parameter exponential family and if the mean life is equal to the mean residual life, then the distribution is exponential.

Finally, there exist many life testing situations which can be best described as mixtures of distributions. In Section 5 we consider the general MRLF and develop a method of obtaining the mixing distribution when the original distribution is exponential. Some examples are discussed, in one of which Morrison's (1978) result "The gamma distribution is the unique mixing distribution of the exponential that leads to a linearly increasing MRLF" is obtained as a special case.

2. PRELIMINARY RESULTS

Let X be a nonnegative random variable denoting the life of a component having distribution function $F(x)$ and probability density function (pdf) $f(x)$. Then the failure rate of X is defined as $\lambda(t) = f(t)/\bar{F}(t)$, where $\bar{F}(t) = 1 - F(t)$ is called the survival function. If $r(t) = E(X - t | X > t)$ denotes the MRLF, then

$$r(t) = \int_t^{\infty} \bar{F}(x) dx / \bar{F}(t) = [\mu - \int_0^t \bar{F}(x) dx] / \bar{F}(t). \quad (1)$$

Differentiating equation (1), one obtains

$$\lambda(t) = [1 + r'(t)] / r(t), \quad (2)$$

which expresses $\lambda(t)$ in terms of $r(t)$.

It is well known that $\lambda(t)$ determines the distribution function uniquely and hence $r(t)$ also determines $\bar{F}(x)$ and we have

$$\bar{F}(t) = \frac{r(0)}{r(t)} \exp\left\{-\int_0^t \frac{dx}{r(x)}\right\} \quad (3)$$

(see Muth, 1977, and Laurent, 1974). This result has also been obtained by Gupta (1979) using a completely different approach.

Thus $\bar{F}(t)$, $\lambda(t)$ and $r(t)$ are all equivalent in the sense that, given one of them, the other two can be determined. Hence in the analysis of survival data, one sometimes estimates $\lambda(t)$ or $r(t)$ instead of $\bar{F}(t)$ according to the convenience of the procedures available.

3. COMPARISON OF MRLF AND THE FAILURE RATE

We say that X has increasing failure rate (IFR) if $\lambda(t)$ is nondecreasing and likewise we define decreasing failure rate (DFR) distributions. The following result shows that IFR implies decreasing MRLF, but the converse is not true. A dual result holds for DFR distributions.

Theorem 1. If F is IFR then the MRLF is decreasing, but not conversely. In particular, $r(t) \leq r(0)$ for IFR distributions.

Proof. The hypothesis implies that $\bar{F}(t+s)/\bar{F}(t)$ is decreasing in t for each $s \geq 0$. [Take logarithmic derivative with respect

to t .] The result follows upon integrating this ratio from $s = 0$ to $s = \infty$. Bryson and Siddiqui (1969) have given an example showing that the converse is not true. Another counter-example to the converse is provided by taking

$$r(t) = 1/(1+t^2) \quad \text{and} \quad \lambda(t) = 1 + t^2 - 2t/(1+t^2).$$

4. INTERPRETATION OF MRLF IN RENEWAL THEORY AND SOME CHARACTERIZATIONS OF THE EXPONENTIAL DISTRIBUTION

Suppose a component operating in a system is replaced upon failure by another component possessing the same life distribution, so that the sequence of component life lengths forms a renewal process. At any time t , the component in operation is identified for study. Let U_t be the age of the component in use at t and let V_t be the remaining life of the component (time from t until failure). The quantities U_t and V_t are known as backward and forward recurrence times, respectively, in renewal theory.

It may be observed that U_t and V_t are independent if and only if the renewal distribution is exponential. Attention has been drawn in the literature for characterizing the Poisson process and hence the exponential distribution by certain properties of the distribution of V_t (see Chung, 1972). Cinlar and Jagers (1973) and Holmes (1974) present a characterization by the mean value of V_t . More specifically, if $E(V_t)$ is finite and is independent of t , then the process is Poisson.

For large values of t , assuming the life distribution is nonlattice, the limiting pdf of U_t or V_t is given by

$$f_1(y) = \bar{F}(y)/\mu, \quad (4)$$

where $\bar{F}(x) = 1 - F(x)$, $\mu = E(X) < \infty$ and $F(0) = 0$ (see Cox, 1962). Let Y be a random variable having pdf (4) representing the residual time of faultless operation of the component. Some applications of the distribution of Y in life length studies are described in Blumenthal (1967) and Scheaffer (1972).

Now $\lambda_1(t)$, the failure rate of Y , is related to the MRLF of X by

$$\lambda_1(t) = \bar{F}(t) / \int_t^\infty \bar{F}(y) dy = 1/r(t).$$

Thus, Theorem 1 is equivalent to the statement "Y is IFR whenever X is IFR." We note that

$$\bar{F}_1(t) = \bar{F}(t)[r(t)/r(0)].$$

Assume that X is IFR so that $r(t)/r(0) \leq 1$. Taking expectation of the preceding equation gives $E(Y) \leq E(X)$, with equality if and only if $r(t) = r(0)$. This gives us the following characterization of the exponential distribution.

Theorem 2. Suppose the renewal distribution is IFR (or DFR). Then $E(Y) = E(X)$ if and only if X has an exponential distribution.

The following theorem shows that if we drop the condition of IFR, then the equality of mean values of X and Y characterizes the exponential distribution in the one-parameter exponential family.

Theorem 3. Suppose the renewal distribution belongs to the one parameter exponential family with pdf

$$f(x; \theta) = c(\theta) h(x) e^{\theta x}.$$

If $E(Y) = E(X)$ for all θ in some interval I, then X has an exponential distribution for all θ in I.

Proof. Using Laplace transform technique, it can be seen that the gamma distribution is characterized within the linear exponential family by the property that its coefficient of variation is independent of θ . (Mafoud, 1977, p. 24; Ratnaparkhi, 1981). Further, having a coefficient of variation equal to unity is characteristic of the exponential distribution within the gamma family. Thus it is sufficient to show that $E(X) = E(Y)$ implies $\text{Var}(X) = [E(X)]^2$. But this is evident since, in general,

$$E(Y) = \int y f_1(y) dy = \int y \bar{F}(y) dy / \mu = E(X^2) / 2\mu.$$

5. MRLF OF A MIXTURE

Recently Morrison (1978) has shown that the gamma distribution is the unique mixing distribution of the exponentials that leads to a linearly increasing MRLF. In the following we have considered the general MRLF and have developed a method of obtaining the mixing distribution when the original distribution is exponential. Some examples are discussed, in one of which Morrison's result is obtained as a special case.

Let X be a nonnegative random variable, denoting the life of a component, having distribution function $F(t; \theta)$ and let $\bar{F}(t)$ be the distribution function of X after mixing on θ . Then

$$E_{\theta}[\bar{F}(t; \theta)] = \bar{F}(t). \quad (5)$$

Suppose now the original distribution is exponential, i.e., $\bar{F}(t; \theta) = \exp(-t\theta)$. We obtain from (5) that

$$G(t) = \bar{F}(t), \quad (6)$$

where $G(t)$ is the Laplace transform of θ . In particular, the mixing distribution is uniquely determined by the unconditional survival function of X , i.e. a mixture of exponentials is identifiable.

Equation (6) has several applications. Firstly, using Bernstein's characterization (see Feller, 1971, p. 439) of Laplace transforms, we may obtain necessary and sufficient conditions for a distribution to be a mixture of exponentials. Secondly, expressing \bar{F} in terms of r or λ , the Laplace transform of θ can be written in terms of either of these functions. Some applications of the first type are summarized in the following theorem.

Theorem 4. Let X be a nonnegative random variable with failure rate $\lambda(t)$ and MRLF $r(t)$. Then (a) X is a mixture of exponentials if and only if its survival function is completely monotone. (b) $\lambda(t)$ is completely monotone if and only if X is a mixture of exponentials with infinitely divisible mixing distribution. (c) If $r(t)$ has a completely monotone derivative, then X is a mixture of exponentials with infinitely divisible mixing distribution.

Proof. Part (a) follows from Bernstein's theorem. For (b) use Theorem 1 of Feller (1971, p. 450). Finally, if $r(t)$ has a completely monotone derivative then $1/r(t)$ is completely monotone (Feller, 1971, Criterion 2, p. 441) as is $(1+R'(t))$. Since the product of completely monotone functions is completely monotone, it follows that $\lambda(t)$ is completely monotone (see equation 2).

Since $\lambda_1 = 1/r$, we see by the argument just given that the (limiting) recurrence time distributions are also mixtures of exponentials with infinitely divisible mixing distribution if $r(t)$ has a completely monotone derivative.

We conclude with some examples in which equation (6) is applied to determine the mixing distribution.

Example 1. Let $r(t) = a + bt$. Employing equation (3) and (6) one obtains

$$G(t) = [a/(a+bt)]^k,$$

where $k = 1 + 1/b$, which is the Laplace transform of a gamma distribution. Thus linear MRLF leads to a gamma distribution for θ . Note that this was the main result obtained by Morrison (1978). Also his expression for $G'(t)/G(t)$ needs a minus sign.

Example 2. Let $\lambda(t) = \alpha t^{\alpha-1}$ with $0 < \alpha < 1$. Proceeding as in Example 1 we get $G(t) = \exp(-t^\alpha)$, which is the Laplace transform of a stable distribution.

The author wishes to thank the referees for some useful comments.

REFERENCES

- Balkema, A. A. and DeHann, L. (1974). Residual life at great age. *Annals of Probability*, 2, 792-804.
- Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics*, 9, 205-218.
- Bryson, C. and Siddiqui, M. M. (1969). Some criteria for aging. *Journal of American Statistical Association*, 64, 1472-1483.
- Chung, K. L. (1972). The Poisson process as a renewal process. *Periodica Mathematica Hungarica*, 2, 41-48.
- Cinlar, E. and Jagers, P. (1973). Two mean values which characterize the Poisson process. *Journal of Applied Probability*, 10, 678-681.
- Cox, D. R. (1962). *Renewal Theory*. Methuen, London.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (second edition). Wiley, New York.
- Gupta, R. C. (1975). On characterization of distributions by conditional expectations. *Communications in Statistics*, 4, 99-103.
- Gupta, R. C. (1979). On the characterization of survival distributions in reliability by properties of their renewal densities. *Communications in Statistics A*, 8, 685-697.
- Hollander, W. and Proschan, F. (1975). Tests for the mean residual life. *Biometrika*, 62, 585-593.
- Holmes, P. T. (1974). Another characterization of the Poisson process. *Sankhya A*, 36, 449-450.
- Laurent, A. G. (1974). On characterization of some distributions by truncated properties. *Journal of the American Statistical Association*, 69, 823-827.
- Mahfoud, M. (1977). *Study of weighted distributions and size-biased sampling*. Ph.D. thesis, The Pennsylvania State University.

- Meilijson, I. (1972). Limiting properties of the mean residual life time function. *Annals of Mathematical Statistics*, 43, 354-357.
- Morrison, D. G. (1978). On linearly increasing mean residual life times. *Journal of Applied Probability*, 15, 617-620.
- Muth, E. J. (1977). Reliability models with positive memory derived from the mean residual life function. In *Theory and Applications of Reliability*, C. P. Tsokos and I. N. Shimi, eds. Academic Press, New York.
- Ratnaparkhi, M. V. (1981). On multivariate splitting model and related characterizations of some statistical distributions. In *Statistical Distributions in Scientific Work*, C. Taillie, G. P. Patil, and B. Baldessari, eds. Reidel, Dordrecht-Holland. Vol. 4, pp. 357-364.
- Scheaffer, R. L. (1972). Size biased sampling. *Technometrics*, 14, 635-644.
- Watson, G. S. and Wells, W. T. (1961). On the possibility of improving the mean useful life of items by eliminating those with short lives. *Technometrics*, 3, 281-298.
- Weiss, G. A. and Dishan, M. (1971). Some economics problems related to burn in programs. *IEEE Transactions on Reliability*, R-20, 190-195.

[Received May 1980. Revised September 1980]

IDENTIFIABILITY PROBLEMS IN THE THEORY OF COMPETING AND COMPLEMENTARY RISKS — A SURVEY

ASIT P. BASU

Department of Statistics
University of Missouri
Columbia, Missouri 65211 USA

SUMMARY. In this expository paper the concepts of competing and complementary risks are defined and a survey of recent results in the area is presented. Identifiability of distributions, both univariate and multivariate, useful in reliability and survival analysis is considered.

KEY WORDS. competing risks, complementary risks, identifiability, reliability, distributions of minimum and maximum, series and parallel systems.

1. INTRODUCTION

The problem of identifiability (or of nonidentifiability) arises naturally in a number of physical situations. The problem, in general terms, can be defined as follows. Let U be an observable random variable whose distribution function belongs to a family $F = \{F_\theta: \theta \in \Omega\}$ of distribution functions indexed by a parameter θ . Here θ could be scalar or vector valued.

We shall say θ is nonidentifiable by U if there are distinct parameter values, θ and θ' , such that $F_\theta(u) = F_{\theta'}(u)$ for all u . In the contrary case we shall say θ is identifiable. It may happen that θ itself is nonidentifiable but a function $\gamma(\theta)$ is identifiable in the following sense: For any $\theta, \eta \in \Omega$, $F_\theta(u) = F_\eta(u)$ for all u implies $\gamma(\theta) = \gamma(\eta)$; in this case we may say θ is partially identifiable.

Puri (1970) has surveyed some examples arising in the general literature. The purpose of the present paper is to present a survey of results available in the area of competing risks and complementary risks. We primarily consider the cases where we have an underlying parametric model. In Sections 2 and 3 we define the problems of competing and complementary risks and discuss the case when the underlying random variables are independently distributed. The case of dependent random variables is considered in Section 4. The problem of estimation is briefly discussed in Section 5. Finally, in Section 6, some open problems and other related areas of research are pointed out.

2. INDEPENDENT RANDOM VARIABLES WITH IDENTIFIED EXTREMUM

The problem of *competing risks*, in its simplest form, may be described as follows: Let X_i be a random variable with distribution function $F_i(x)$, ($i = 1, 2, \dots, p$). We assume that the X_i 's are not observable, but that $U = \min(X_1, \dots, X_p)$ is. We would like to estimate the F_i 's given the observations on U .

This model finds interesting applications in a number of fields but particularly in problems of survival analysis and reliability theory. Thus in problems of competing risks, or in studying reliability of complex systems, an individual or a system of components may be exposed to p different causes of death (failure) where X_i is the time to death from the i th cause. Although one would like to know about the distribution of the X_i 's, only observations on U 's will be available.

Basu and Ghosh (1980) give some of these examples. For other examples and a survey of the area see Birnbaum (1979) and David and Moeschberger (1978).

When the X_i 's are not identically distributed, the problem of identifiability is illustrated by the following example.

Example 1. Let X_i ($i = 1, 2, 3, 4$) be independent random variables and suppose X_i is exponentially distributed with $E(X_i) = 1/\lambda_i$. If $\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4$, then $\min(X_1, X_2)$ and $\min(X_3, X_4)$ are identically distributed.

It is possible to introduce additional random variables so that the enhanced family with the added information becomes identifiable. In this case we call the original family of

distributions *rectifiable*. Let I be an integer-valued random variable ($I = 1, 2, \dots, p$). (U, I) is called an *identified minimum* if $I = k$ when $U = \min(X_1, X_2, \dots, X_p) = X_k$. In the absence of I , U will be called a *non-identified minimum*.

Basu and Ghosh (1980) consider a dual problem, called the problem of *complementary risks*, where instead of observing the minimum (identified or non-identified) one observes the maximum $V = \max(X_1, X_2, \dots, X_p)$. Basu and Ghosh describe examples showing that this problem also occurs naturally in survival analysis and reliability theory.

We now consider the case when the X_i 's are independently but not identically distributed. The joint probability distribution of (U, I) is specified by the monotonic functions

$$H_k(x) = P(U \leq x, I = k), \quad k = 1, 2, \dots, p.$$

Then Berman (1963) has obtained the following theorem.

Theorem 1. The set of functions $\{H_k(x)\}$ is related to the set $\{F_j(x)\}$ by the functional equations

$$H_k(x) = \int_{-\infty}^x \prod_{j \neq k} [1 - F_j(t)] dF_k(t), \quad k = 1, 2, \dots, p.$$

This solution of this set of equations is

$$F_k(x) = 1 - \exp\left\{- \int_{-\infty}^x \left[1 - \sum_{j=1}^p H_j(t)\right]^{-1} dH_k(t)\right\},$$

where $k = 1, 2, \dots, p$.

Since the minimum U and the maximum V satisfy the relation $\max(X_1, X_2, \dots, X_p) = -\min(-X_1, -X_2, \dots, -X_p)$, it follows that similar results also hold for the maximum.

The next natural questions are: (a) What if the minimum (maximum) is not identified? Can we obtain the distributions of the X_k from that of U ? And (b) What can we say about the identifiability of the F_k 's if the X_k 's are not independently distributed? We will discuss these next.

3. INDEPENDENT RANDOM VARIABLES WITH NONIDENTIFIED EXTREMUM

In case the extremum U is not identified, one can still uniquely determine $F_k(x)$ under certain conditions. To this end Basu and Ghosh (1980) obtain the following theorem.

Theorem 2. Let F be a family of pdf on R_1 with support (a, b) which are continuous and are positive to the left of some point A and such that if f and g are any two distinct members of F then $\lim\{f(x)/g(x)\}$ as $x \rightarrow a$ exists and equals either 0 or ∞ . Let X_1, \dots, X_p be independent random variables with respective pdf's f_1, f_2, \dots, f_p in F and let Y_1, Y_2, \dots, Y_2 be independent random variables with respective pdf's belonging to F . If $\min(X_1, \dots, X_p)$ and $\min(Y_1, \dots, Y_q)$ have identical distributions, then $p = q$ and there exists a permutation (k_1, k_2, \dots, k_p) of $(1, 2, \dots, p)$ such that the pdf of Y_i is f_{k_i} , $(i = 1, 2, \dots, p)$.

Anderson and Ghurye (1977) proved a similar theorem for the maximum. As an application of the above theorem, consider the following examples.

Example 2. F is the family of normal distributions

$$\phi(x|\mu, \sigma) = \exp[-(x-\mu)^2/2\sigma^2]/\sqrt{(2\pi\sigma^2)}.$$

Now,

$$\lim_{x \rightarrow -\infty} \frac{\phi(x|\mu_2, \sigma_2)}{\phi(x|\mu_1, \sigma_1)} = \begin{cases} 1, & \text{if } \mu_1 = \mu_2 \text{ and } \sigma_1 = \sigma_2 \\ 0, & \text{if } \sigma_2 < \sigma_1, \text{ or } \sigma_1 = \sigma_2 \text{ and } \mu_2 > \mu_1 \\ \infty, & \text{if } \sigma_2 > \sigma_1, \text{ or } \sigma_1 = \sigma_2 \text{ and } \mu_2 < \mu_1. \end{cases}$$

Conditions of the above theorem are met. Hence the distributions are identifiable.

Example 3. F is the family of exponential distributions

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda > 0.$$

$$\text{Here } \lim_{x \rightarrow 0} \frac{f_{\lambda_2}(x)}{f_{\lambda_1}(x)} = \begin{cases} 1, & \text{if } \lambda_1 = \lambda_2 \\ \lambda_2/\lambda_1, & \text{if } \lambda_1 \neq \lambda_2 \end{cases}$$

The conditions of the above theorem are not met. By Example 1, the distributions are not identifiable. (Note, however, if the maximum is observed both normal and exponential distributions are identifiable.)

There may, however, be situations when the conditions of the above theorem are not met and yet the underlying family of distributions is identifiable. For example, Basu and Ghosh (1980) have proved the following theorems.

Theorem 3. Let X_i have a gamma distribution with index α_i and scale parameter β_i , ($i = 1, 2, 3, 4$). Assume that α_1 and α_2 are not both equal to one and α_3 and α_4 are not both equal to one. Let X_1 and X_2 be independent and let X_3 and X_4 be independent. If the distribution of $\min(X_1, X_2)$ is identical with that of $\min(X_3, X_4)$, then either

$$(\alpha_1, \alpha_2) = (\alpha_3, \alpha_4) \quad \text{and} \quad (\beta_1, \beta_2) = (\beta_3, \beta_4)$$

or
$$(\alpha_1, \alpha_2) = (\alpha_4, \alpha_3) \quad \text{and} \quad (\beta_1, \beta_2) = (\beta_4, \beta_3).$$

Theorem 4. Let $X_i \sim W(p_i, \theta_i)$, ($i = 1, 2, 3, 4$) be independent Weibull random variables. If the distribution of $\min(X_1, X_2)$ is the same as that of $\min(X_3, X_4)$, then either

$$(p_1, \theta_1) = (p_3, \theta_3) \quad \text{and} \quad (p_2, \theta_2) = (p_4, \theta_4),$$

or
$$(p_1, \theta_1) = (p_4, \theta_4) \quad \text{and} \quad (p_2, \theta_2) = (p_3, \theta_3)$$

provided $p_1 \neq p_2$.

In Theorem 4 we exclude the case $p_1 = p_2$. For if $p_1 = p_2 = p$, say, the problem, after using the transformation $Y_i = X_i^p$, ($i = 1, 2$), reduces to that of the exponential distribution.

4. DEPENDENT RANDOM VARIABLES

We next consider the case when (X_1, X_2, \dots, X_p) are dependent. We restrict our discussion primarily to the case $p = 2$. There are many physical situations where it is desirable to test the assumption that the X_i 's are independent. It is therefore natural to study the extent to which (U, I) or U determines whether the joint distribution of X 's can be identified. To this end Basu and Ghosh (1978) pointed out some difficulties, using the following construction. Let $\bar{F}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$, and $\bar{F}(x_1, x_2) = \partial \bar{F}(x_1, x_2) / \partial x_i$, $(i = 1, 2)$. For simplicity assume that the density of (X_1, X_2) is everywhere positive. Let

$$\bar{G}_i(x) = \exp \left\{ - \int_{-\infty}^x -\bar{F}_i(Z, Z) [F(Z, Z)]^{-1} dZ \right\}$$

and assume $\int_{-\infty}^{\infty} -\bar{F}_i(Z, Z) (\bar{F}(Z, Z))^{-1} dZ$ diverges for $i = 1, 2$.

Then $G_i(x) = 1 - \bar{G}_i(x)$ is a distribution function and (U, I) has the same distribution whether (X_1, X_2) is distributed according to $F(x_1, x_2)$ or according to $G(x_1, x_2) \equiv G_1(x_1) \cdot G_2(x_2)$. Thus the problem could have a satisfactory solution only if F is known to belong to a well-specified parametric family of distributions. Similar results for nonidentifiability, in the absence of specific parametric models, have also been considered by Miller (1977), Tsiatis (1975), and Rose (1973). Tsiatis (1978) further illustrated the magnitude of this problem with some actual data.

No general result on identifiability for dependent random variables is currently available. However Peterson (1975) has obtained some interesting inequalities when $p = 2$.

Bivariate Normal Distribution. In case of specific parametric models, questions of identifiability have been settled for a number of distributions. We summarize some of these results. Let $(X_1, X_2) \sim \text{BVN}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_{12})$. Nádas (1971) showed that, if $0 < \rho_{12} < 1$, the distribution of the identified minimum of a normal pair determines the distribution of the pair. Nádas' proof is not complete, however. Basu and Ghosh (1978) completed Nádas' proof and extended it to the case of nonidentified minimum. Recently Gilliland and Hannan (1980) gave an elegant solution of the problem.

No result for the general case is available as yet. However Basu and Ghosh (1978) established the identifiability of the tri-variate normal distribution given the distribution of the identified minimum and given that, for each pair of random variables X_i, X_j with correlation coefficient ρ_{ij} and deviations σ_i and σ_j ,

$$1 - \rho_{ij}\sigma_i/\sigma_j > 0, \quad i, j = 1, 2, 3; i \neq j.$$

Most of the identification problems considered so far are also valid when the maximum V , instead of the minimum U , is observable. For if \tilde{X} is p variate normal, so is $-\tilde{X}$ and $\max(X_1, \dots, X_p) = -\min(-X_1, \dots, -X_p)$, and thus identification problems for the maximum can be restated in terms of corresponding problems for the minimum. Also, any bivariate distribution obtained through strict monotone transformation of normal variables will be identifiable. The bivariate lognormal distribution is thus identifiable.

Next we consider identifiability of several bivariate distributions useful in reliability theory and survival analysis. In particular, we consider bivariate exponential distributions. A survey of some of these distributions is presented in Basu and Block (1975). These include the bivariate exponential distributions of Marshall and Olkin (1967), Block and Basu (1974), and Gumbel (1960). Basu and Ghosh (1978, 1980) have considered identifiability of these distributions. Their results are summarized below.

(a) *Marshall and Olkin Bivariate Exponential*. The tail probability of this distribution is given by

$$\bar{F}(x_1, x_2) = \exp[-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \max(x_1, x_2)],$$

$$x_1, x_2, \lambda_1, \lambda_2 > 0, \lambda_{12} \geq 0.$$

Here all parameters are identifiable if (U, I) is observed. However, if only U is observed the parameters are not identifiable.

(b) *Block-Basu Model*. Here the joint density function is given by

$$\{\lambda\lambda_1(\lambda_2 + \lambda_{12})/(\lambda_1 + \lambda_2)\} \exp\{-\lambda_1 x_1 - (\lambda_2 + \lambda_{12})x_2\} \quad \text{if } x_1 < x_2$$

$$\text{and } \{\lambda\lambda_2(\lambda_1 + \lambda_{12})/(\lambda_1 + \lambda_2)\} \exp\{-(\lambda_1 + \lambda_{12})x_1 - \lambda_2 x_2\} \quad \text{if } x_2 > x_1$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. Here the parameters are not identifiable at all. Note that the model proposed by Freund (1961) is also not identifiable, since it is related to the Block-Basu model.

Because of the underlying physical assumptions neither Marshall-Olkin nor Block-Basu is considered a suitable physical model when the maximum is observed.

(c) *Gumbel Model I.* Gumbel (1960) proposed two bivariate exponential distributions. The first is given by

$$F(x_1, x_2) = 1 - \exp(-\lambda_1 x_1) - \exp(-\lambda_2 x_2) \\ + \exp(-\lambda x_1 - \lambda_2 x_2 - \lambda_2 x_2 - \lambda_{12} x_1 x_2),$$

where $x_1, x_2, \lambda_1, \lambda_2 > 0$, $\lambda_{12} \geq 0$. Here, the parameters are identifiable if (U, I) is observed. However, if the nonidentified minimum U is observed only λ_{12} and $\lambda_1 + \lambda_2$ are identifiable. If the nonidentified maximum $V = \max(x_1, x_2)$ is observable, then λ_{12} is identifiable and $(\lambda_1, \lambda_{12})$ is identifiable up to a permutation.

(d) *Gumbel Model II.* Here the distribution function is given by

$$[1 - \exp(-\lambda_1 x_1)][(1 - \exp(-\lambda_2 x_2))[1 + \lambda_{12} \exp(-\lambda_1 x_1 - \lambda_2 x_2)]].$$

If U is observable, λ_{12} is identifiable and (λ_1, λ_2) is identifiable up to a permutation.

(e) *Bivariate Weibull Distribution.* The bivariate Weibull distribution can be defined by its survival function

$$\bar{F}(x_1, x_2) = \exp \left[-\lambda_1 x_1^{p_1} - \lambda_2 x_2^{p_2} - \lambda_{12} \max(x_1^{p_1}, x_2^{p_2}) \right].$$

If U is observable, p_1 and p_2 are identifiable up to permutation. Also λ_1 and $\lambda_2 + \lambda_{12}$, or λ_2 and $\lambda_1 + \lambda_{12}$ are identifiable.

5. ESTIMATION OF PARAMETERS

Estimation of parameters based on (U, I) , the identified minimum, has been considered extensively. For the bivariate

normal distribution, Basu and Ghosh (1978) have considered the estimation based on U alone. Similar results can be obtained using V .

For the general bivariate model the pdf of V , assuming independence, is given by

$$f_V(t) = f_1(t)F_2(t) + f_2(t)F_1(t).$$

The parameters can therefore be estimated numerically using the method of maximum likelihood. The method of moments may provide a simpler technique.

6. CONCLUDING REMARKS

In this section we point out a number of areas in which additional work is being carried out and we also point out a few open problems.

So far we have assumed, in the competing risks problem, that the observed random variable U is the minimum of p (unobserved) random variables X_1, X_2, \dots, X_p . In analyzing mortality data the above is interpreted as follows. Consider a population in which p causes of death, C_1, C_2, \dots, C_p , are operating. Each individual in this population is exposed to the risk of dying from any one of these causes. One can recognize two kinds of distributions associated with death due to cause C_i : (a) the survival distribution $\bar{F}_{ia}(t)$ due to cause C_i , conditionally that C_i is the cause of death, *in the presence* of other causes; and (b) the survival distribution $\bar{F}_i(t)$ due to C_i , if C_i is acting alone.

It is tacitly assumed that corresponding forces of mortality (failure rates) are the same. Gail (1975), Elandt-Johnson (1976), and others have explored the implication of this assumption.

A second assumption is that the potential survival times X_i 's are independently distributed with continuous distribution function F_i , ($i = 1, 2, \dots, p$). Some results to this end are given in Sections 2, 3, and 4. For another interesting direction of research see Miller (1977), Desu and Narula (1977), Langberg, Proschan, and Quinzi (1977, 1978), and the references therein. Desu and Narula consider the problem of estimating the distribution function $F_i(t) = P(X_i \leq t)$ and provide a sufficient

condition on the distribution of (X_1, \dots, X_p) under which such an estimation is possible.

A third direction of research has been the interpretation of competing risks problems in terms of some stochastic process. Chiang (1968) has studied the problem of competing risks using time-nonhomogeneous Markov processes. Clifford (1977) and Berlin, Brodsky, and Clifford (1977) have considered the problem of identifiability for this situation.

The problems described in the previous sections can be extended in several directions. The author is currently working on some of the problems stated below.

(a) As mentioned before, Theorem 2 of Section 3 does not provide a necessary and sufficient condition. It would be desirable to improve on this result.

(b) Most of the identifications results obtained so far apply to the case of two competing causes. These need to be generalized to the case of any number of variables. Some results to this end have been obtained by Basu and Ghosh (1980b). Algorithms for estimating the parameters should also be obtained.

(c) The concept of competing risks is well known. Basu and Ghosh (1980) coined the term *complementary risks* for the dual problem. In reliability theory the corresponding problems are for series and parallel systems. It is natural to pose the following general problem corresponding to a k -out-of- p system ($k \leq p$). Recall a system is called k -out-of- p if the system operates so long as k or more components function. Let $X_{(r)}$ be the r th order statistic among X_1, X_2, \dots, X_p . Suppose only $X_{(r)}$ is observable, where $r = p - k + 1$. We assume the X 's are independent.

Given the distribution of some $X_{(r)}$, can we uniquely determine the distribution of each X_i , ($i = 1, 2, \dots, p$)? If $r = 1$, we obtain the case of competing risks, and if $r = p$, the case of complementary risks. For identically distributed X_i 's, specification of the distribution of any order statistic is known to completely determine the common distribution of the X 's (cf. Galambos, 1975, p. 79). Therefore we turn to the case in which the X 's are not equal in distribution, and we consider the r th identified order statistic $(X_{(r)}, I)$ where $X_{(r)} = X_k$ when $I = k$. By Berman's theorem (cf. Section 2), the distribution of the X 's are determined by the distribution of the identified

minimum (or maximum). For exponentially distributed variates we have the following general result

Theorem 5. Let X_1, X_2, \dots, X_p be independent and exponentially distributed with $E(X_i) = 1/\lambda_i$. Then, for any r , the joint distribution of the r th identified order statistic uniquely determines the values of $\lambda_1, \lambda_2, \dots, \lambda_p$.

Proof. Let X_i have density f_i and survival function \bar{F}_i and consider the joint density function of $(X_{(r)}, I)$, i.e.,

$$\frac{d}{dt} P(X_{(r)} \leq t, I = k), \quad k = 1, 2, \dots, p. \quad (1)$$

For each k , this is assumed to be a known function of t . For notational simplicity, take $p = 3$ and $r = 2$. With $k = 1$, say, the joint density (1) becomes

$$\begin{aligned} f_1(t) [F_2(t)\bar{F}_3(t) + \bar{F}_2(t)F_3(t)] \\ = \lambda_1 e^{-\lambda_1 t} [e^{-\lambda_2 t} + e^{-\lambda_3 t} - 2e^{-(\lambda_2 + \lambda_3)t}]. \end{aligned} \quad (2)$$

Since distinct exponential functions are linearly independent, $-\lambda_1$ is uniquely determined as the coefficient of the most rapidly decreasing exponential in (2). Similarly for λ_2 and λ_3 . The proof for general p and r is analogous: λ_k is a known multiple of the most rapidly decreasing exponential in the k th section of (1).

The preceding argument yields a slightly more general result: specification of the k th section of (1) determines λ_k uniquely and determines the other λ 's up to a permutation. Further results along these lines have been obtained by Basu and Ghosh (1980b).

ACKNOWLEDGEMENT

The paper was prepared under contract N00014-78-C-0655 for the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

REFERENCES

- Anderson, T. W. and Ghurye, S. G. (1977). Identification of parameters by the distribution of a maximum variable. *Journal of the Royal Statistical Society, Series B*, 39, 337-342.
- Basu, A. P. and Block, H. W. (1975). On characterizing univariate and multivariate exponential distributions with applications. In *Statistical Distributions in Scientific Work, Vol. 3*, G. P. Patil, S. Kotz, and J. K. Ord, eds. Reidel, Dordrecht-Holland. pages 399-421.
- Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis*, 8, 413-429.
- Basu, A. P. and Ghosh, J. K. (1980a). Identifiability of distributions under competing risks and complementary risks model. *Communications in Statistics* (to appear).
- Basu, A. P. and Ghosh, J. K. (1980b). Unpublished manuscript.
- Berlin, B., Brodsky, J., and Clifford, P. (1977). Testing disease dependence in survival experiments with serial sacrifice. (Unpublished.)
- Berman, S. M. (1963). Notes on extreme values, competing risks, and semi-Markov processes. *Annals of Mathematical Statistics*, 34, 1104-1106.
- Birnbaum, Z. W. (1979). On the mathematics of competing risks. DHEW Publication Number: (PHS) 79-1351, National Center for Health Statistics, Hyattsville, Maryland.
- Block, H. W. and Basu, A. P. (1974). A continuous exponential extension. *Journal of the American Statistical Association*, 69, 1031-1037.
- Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. Wiley, New York.
- Clifford, P. (1977). Nonidentifiability in stochastic models of illness and death. *Proceedings of the National Academy of Sciences, USA*, 74, 1338-1340.
- David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks*. Griffin, London.
- Desu, M. M. and Narula, S. C. (1977). Reliability estimation under competing causes of failure. *The Theory and Applications of Reliability, Vol. 2*, C. P. Tsokos and I. N. Shimi, eds. Academic Press, New York. Pages 471-481.
- Elandt-Johnson, R. (1976). Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. *Scandinavian Actuarial Journal*, 1, 37-51.
- Freund, J. E. (1961). A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56, 971-977.
- Gail, M. (1975). A review and critique of some models used in competing risks analysis. *Biometrics*, 21, 209-222.

- Galambos, J. (1975). Characterizations of probability distributions by properties of order statistics, I. In *Statistical Distributions in Scientific Work*, Vol. 3, G. P. Patil, S. Kotz, and J. K. Ord, eds. Reidel, Dordrecht-Holland, Pages 71-88.
- Gilliland, D. C. and Hannan, J. (1980). Identification of the ordered bivariate normal distribution by minimum variate. To be published.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698-707.
- Langberg, N., Proschan, F., Quinzi, A. J. (1977). Transformations yielding reliability models based on independent random variables: A survey. In *Applications of Statistics*, P. R. Krishnaiah, ed. North Holland Publishing Company, Amsterdam. Pages 323-337.
- Langberg, N., Proschan, F., Quinzi, A. J. (1977). Converting dependent models into independent ones, preserving essential features. *Annals of Probability*, 6, 174-181.
- Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.
- Miller, D. R. (1977). A note on independence of multivariate lifetimes in competing risks model. *Annals of Statistics*, 5, 576-579.
- Nádas, A. (1970). On estimating the distribution of a random vector when only the smallest co-ordinate is observable. *Technometrics*, 12, 923-924.
- Nádas, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics*, 13, 201-204.
- Peterson, A. V., Jr. (1975). Bounds for a joint distribution function with fixed sub-distribution functions; application to competing risks. Technical Report Number 7, Stanford University, Division of Biostatistics.
- Puri, P. S. (1979). On certain problems involving nonidentifiability of distributions arising in stochastic modeling. In *Optimizing Methods in Statistics*, J. S. Rustagi, ed. Academic Press, New York. Pages 403-417.
- Rose, D. M. (1973). An investigation of dependent competing risks. Unpublished doctoral thesis, University of Washington.
- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, USA*, 72, 20-22.
- Tsiatis, A. (1978). An example of nonidentifiability in competing risks. *Scandinavian Actuarial Journal*, 3, 235-239.

[Received June 1980. Revised September 1980]

DEPENDENCE CONCEPTS FOR STOCHASTIC PROCESSES

DENNIS S. FRIDAY

Statistical Engineering Laboratory
National Bureau of Standards
Boulder, Colorado 80303 USA

SUMMARY. Dependence concepts are a relatively recent development in multivariate distribution theory. They allow dependence to be incorporated into a problem without requiring specific model assumptions, and define classes of multivariate distributions with useful properties. It is not apparent that related notions have also evolved in the theory and applications of stochastic processes. Multivariate concepts and stochastic process concepts, however, have had no influence on each other. In this paper an overview is presented of the work in stochastic processes that is analogous to multivariate dependence concepts.

KEY WORDS. Dependence concepts, fractional Brownian motion, mixing, power-law spectra, self-similarity, stochastic processes.

1. INTRODUCTION

Stochastic independence may be thought of as a singularity in the continuum of all possible stochastic relationships among a set of random variables. It is an assumption that pervades both theoretical and applied statistical work. The reasons for this perhaps excessive use of independence may be attributed to two complementary factors: (i) The mathematical tractability of statistical problems under the assumption of independence, and consequently, the availability of the numerous known results. (ii) The comparative void that confronts a statistician who attempts to introduce dependence into a problem without making very restrictive assumptions about its form.

While new results are continually being published that make the first factor even more attractive, much recent work on dependence concepts for multivariate distributions is generally directed toward neutralizing the second factor. Considerable progress has been made in a comparatively short time span, much of it oriented toward reliability applications. Dependence concepts that have been developed include: Association, Positive Quadrant Dependence, Positive and Negative Dependence, and many others. See Ahmed *et al.*, (1978), Barlow and Proschan (1975), Block *et al.*, (1980), Esary and Proschan (1972), Kimeldorf and Sampson (1978), and Lehmann (1966).

This work on dependence concepts has not been linked with general problems in stochastic processes even though finite dimensional distributions are central in the study of the latter. Special cases have been considered [Ross (1979), Esary *et al.*, (1967)]. It is also not apparent that related notions have evolved within the field of stochastic processes. We show that such work does exist and that general concepts of dependence are important for both theory and applications. The objective of this paper is to present a firm, heuristic understanding of the more important concepts without all of the mathematical details. The theorems and their proofs and many perturbations thereof exist in the literature. The overview we hope to present does not.

2. GENERAL PROPERTIES OF RANDOM PROCESSES

A few relevant definitions and preliminaries are in order. Univariate, real valued stochastic processes will be of primary interest here. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let T be the set of indices of the process. Then the finite, real valued function $X(t, \omega)$ is a stochastic process if it is a measurable function of $\omega \in \Omega$ for each $t \in T$. We will suppress the argument ω and write $X(t)$. In most cases, T will be either the real line or the (doubly infinite) integers. For any fixed $t \in T$, $X(t)$ is a random variable with distribution function $F(x; t) = \mu(\{X(t) \leq x\})$. Similarly a finite set $\{t_1, \dots, t_n\}$; $t_i \in T$, defines a random vector $X(t_1), \dots, X(t_n)$ with the multivariate distribution function

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = \mu(\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}).$$

Any distribution of this type, derived from the $X(t)$ process, is called a finite dimensional distribution. The family of all of its finite dimensional distributions uniquely defines a stochastic process [Kolmogorov (1940)].

Assume that the finite dimensional distributions have finite first and second order moments. The mean function of a stochastic process $X(t)$ is given by $m(t) = E[X(t)]$, $t \in T$. It is a function of t , and when it is convenient the zero mean process $X(t) - m(t)$ will be used instead of $X(t)$. The covariance function of $X(t)$ (zero mean) is given by $C(s, t) = E[X(s)X(t)]$; $s, t \in T$. It has the following properties: *Symmetry*: $C(s, t) = C(t, s)$;

Non-negativity: $C(t, t) = E[x(t)^2] > 0$; *Schwartz Inequality*:

$|C(s, t)|^2 < E[X(s)]^2 E[X(t)]^2$; and *Nonnegative Definiteness*:

$$\sum_{i,j=1}^n C(t_i, t_j) \alpha_i \alpha_j \geq 0 \quad \text{for all } \{t_1, \dots, t_n\}; t_i \in T$$

and all $\alpha_1, \dots, \alpha_n$.

A stochastic process $X(t)$ is strictly stationary if for any t_1, \dots, t_n ; $t_i \in T$, and any s such that $t_i + s \in T$ the random vectors $X(t_1), \dots, X(t_n)$ and $X(t_1 + s), \dots, X(t_n + s)$ have the same finite dimensional distributions. Any finite dimensional distribution of a strictly stationary process is invariant under a translation in time. A weaker definition is often used. A stochastic process $X(t)$ is stationary if $E[X(t)] = m$ and $E[X(s)X(t)] = C(t-s)$. A stationary process has a constant mean value and the covariance between the random variables defined at any two time points is a function only of the distance between these points. An important aspect of stochastic processes is their spectral representations. Let $X(t)$; $t \in (-\infty, \infty)$ be a real valued stationary stochastic process with covariance function $C(\tau)$. Then $C(\tau)$ has the representation:

$$C(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda} dS(\lambda); -\infty < \tau < \infty,$$

where $S(\infty) - S(-\infty) = C(0) < \infty$ and $S(\lambda)$ is real, nondecreasing and bounded. The process $X(t)$ is said to have spectral distribution function $S(\lambda)$. $S(\lambda)$ denotes the relative magnitude of harmonic components of frequency λ present in $X(t)$. The spectrum represents the distribution of the total variance $C(0)$ of the process $X(t)$ over frequency. The region where $|\lambda|$ is small is called the low frequency region and large $|\lambda|$ corresponds to the high frequency region. Note that the set in which λ varies depends on whether X is continuous or discrete. In the equispaced, distance Δ , discrete time case $C(\tau)$ has spectral representation for $\tau = 0, \pm \Delta, \pm 2\Delta, \dots$

$$C(\tau) = \int_{-\pi/\Delta}^{+\pi/\Delta} e^{i\tau\lambda} d\left\{ \sum_{j=-\infty}^{\infty} S(\lambda + 2\pi j/\Delta) - S((2\pi j - \pi)/\Delta) \right\},$$

where the domain of the spectral density is now $\lambda \in [-\pi/\Delta, +\pi/\Delta]$.

The covariance and mean functions are clearly related to the serial dependence in a stochastic process. Deterministic dependencies, such as $m(t)$, are simple in nature and not of concern in this study. The covariance function, however, possesses non-trivial information related to dependence in a stochastic process. Given a stochastic process with covariance function $C(\tau)$, the most direct link between the process and multivariate dependence concepts is to choose a set of time points t_1, \dots, t_n and apply the dependence concepts to $X(t_1), \dots, X(t_n)$. We will digress, however, and consider certain aspects of dependence that have evolved solely within the theory and applications of stochastic processes.

3. ASYMPTOTIC INDEPENDENCE CONDITIONS

One class of dependence considerations originating in the theory of stochastic processes is asymptotic in nature and has no counterpart in multivariate distribution theory. Historically, ergodic theory and central limit theory for dependent sequences were the motivation for these concepts. Ergodicity is of fundamental importance for statistical applications involving stochastic processes. An ergodic process is one for which the time average and the ensemble average are equivalent. The practical implication is that the single "long-term" realization of an ergodic stochastic process may be used to infer properties of the family of all possible realizations of the process. That is:

$$\lim_{n \rightarrow \infty} 1/n \sum_{t=0}^n X(t; \omega) = E_{\omega}[X(t; \omega)].$$

Ergodic theory had its origins in Statistical Mechanics [Hopf (1937), Kac (1959)] and applications to probability followed [Rosenblatt (1956), Blum and Rosenblatt (1956), Rosenblatt (1961)].

Let $\dots, X_{-1}, X_0, X_{+1}, \dots$ be a strictly stationary stochastic process with $E(X_1) < \infty$ and let L_r denote an event in the smallest σ -field generated by events on $\{\dots, X_{r-1}, X_r\}$ and let R_s denote an event in the corresponding σ -field on $\{X_s, X_{s+1}, \dots\}$. Then a necessary and sufficient condition for ergodicity is that for all L_r and R_s , $r, s = \dots, -1, 0, +1, \dots; r < s$

$$\lim_{n \rightarrow \infty} 1/n \sum_{s-r=1}^n P(L_r R_s) = P(L_r)P(R_s).$$

While the details are covered in various ways by many authors, the basic principle is that any events invariant under shift transformations must differ from the sample space Ω or the null set by, at most, measure zero. This property is also known as metric transitivity.

Ergodicity, however, is not a strong enough condition for certain limit theorems. One stronger condition is called the mixing condition. Let classes of events L_r and R_s be as defined above. Then the process $\dots, X_{-1}, X_0, X_{+1}, \dots$ is mixing if

$$\lim_{s-r \rightarrow \infty} P(L_r R_s) = P(L_r)P(R_s).$$

Therefore, mixing means that the information between any two events decreases as they become separated in time (or space, etc.). More precisely, any two events become stochastically independent as the regions of the process, on which they are defined, move apart. It is also implicit in this definition that the events cannot be made large enough to offset this decreasing dependence.

For certain applications the mixing condition is too weak. Zhurbenko (1975), for example, shows that it is not sufficient for certain time series results. A stronger definition of asymptotic independence is the following. Let L_r and R_s be as defined previously and let the function $g(x) \rightarrow 0$ as $x \rightarrow \infty$. Then the process $\{\dots, X_{-1}, X_0, X_1, \dots\}$ is said to be g-mixing if for all $L_r, R_s, r < s$,

$$|P(L_r R_s) - P(L_r)P(R_s)| \leq g(s-r).$$

Clearly, all g-mixing processes are mixing processes that have additional constraints on the rate at which independence is achieved. A comment on terminology is in order, since authors are not consistent. Our g-mixing is occasionally called uniform mixing or strong mixing, while other authors use the term strong mixing to describe our mixing, and the term weak mixing to describe the mixing condition which is sufficient for ergodicity. Others, in particular the Russian literature, use the term regularity. Many variations of these conditions can be found in the literature [Kolmogorov and Rozanov (1960)],

4. LONG-TERM DEPENDENCE

This section deals with mathematical descriptions of dependence which were originally motivated by observations of real world phenomena. While the previous section dealt with short-term dependence, we now consider processes for which there is a strong dependence even for large time lags.

Hydrology of the Nile river was the stimulus for much of the work that we will discuss. Data on water flow in the Nile reveals that the next year, and many subsequent years are very likely to be similar to the current year. Years of drought are followed by years of drought, and years of flood are followed by years of flood. H. E. Hurst, an English physicist, recognized the phenomena and developed statistical procedures to deal with a seemingly purely hydrological problem. B. B. Mandelbrot showed that such phenomena are far more general than hydrology and has contributed much to their mathematical description [Lawrance and Kottegoda (1977)]. Mandelbrot (1977) has developed convincing arguments that geographical coastlines, music, mountainous skylines, noise in electronic devices, and turbulence have much in common stochastically with river flow.

This discussion of dependence begins with a statistical procedure that provides a measure of the dependence of a stochastic process [Mandelbrot and Taqqu (1979)].

If $X(t)$; $t \geq 0$ is a continuous time random process let

$$X^*(t) = \int_0^t X(s)ds, \quad X^{2*}(t) = \int_0^t X^2(s)ds, \quad \text{and} \quad X^{*2} = (X^*)^2.$$

If $X(t)$; $t = 1, 2, \dots$ is a discrete time process then let

$$X^*(0) = 0, \quad X^*(t) = \sum_{s=1}^t X(s), \quad \text{and} \quad X^{2*}(t) = \sum_{s=1}^t X^2(s).$$

Now define for $d > 0$: $\Delta(u) = X^*(u) - u/d X^*(d)$,

$$R(d) = \sup_{0 \leq u \leq d} \Delta(u) - \inf_{0 \leq u \leq d} \Delta(u), \quad \text{and} \quad S^2(d) = 1/d X^{2*}(d) - 1/d^2 X^{*2}(d)$$

Finally, $Q(d) = R(d)/S(d)$; $d > 0$. $R(d)$ is the adjusted range, $S(d)$ the sample standard deviation, and $Q(d)$ is the rescaled adjusted range. $Q(d)$ is also called the R/S statistic. The lag d is the right end point of the time interval $[0, d]$ considered. These definitions may be generalized to $Q(t, d)$ by considering the interval, $[t, t+d]$.

Observe that $Q(d)$ is invariant under linear transformations on $X(t)$. That is, $Q(d)$ has the same value for a sample of $X(t)$ as for $aX(t) + b$. Furthermore, $R(d)$ and $S(d)$ are each

invariant if $a=1$, that is, under level shifts of $X(t)$. The original derivation of $Q(d)$ was heuristic and based on hydrological considerations. The objective was to determine the size of reservoir that, based upon some interval of past history of a river, would have contained all floods and for which the reservoir would never have run dry during $[0, d]$. The reservoir must then contain an initial quantity of $-\inf \Delta(u)$. It is assumed that the water is flowing out of the reservoir at a constant rate so that the amount withdrawn, during time $[0, d]$, is equivalent to the total discharge $X^*(d)$ into the reservoir. This information is very important, for example, in designing a dam. Hurst added the denominator $S(d)$ as a normalizing factor. Mandelbrot and Wallis (1969) discovered that this was a very fortuitous choice. Among other properties they showed that $Q(d)$ is robust against deviations from normality in $X(t)$ and against $E[X^2(t)] = \infty$, infinite variance.

The purpose of the $Q(d)$ statistic is to identify whether or not the process from which a given set of data arose exhibits long-term dependence and, if it does, to characterize the level of dependence by a single number. The symbol J has been assigned to this number by Mandelbrot, who claims that the earliest reference to such phenomena occurs in the biblical story of Joseph. In slightly more detail, if there exists a real number J , such that $d^{-J}Q(d)$ converges in distribution to non-degenerate limits as $d \rightarrow \infty$, then this J is called the asymptotic Hurst Exponent or the R/S exponent. It has been shown that J has range $0 \leq J \leq 1$. Heuristically, if the $Q(d)$ corresponding to a stochastic process can be normalized for some J then the sample $Q(d)$'s will fluctuate about the line d^J . A graphical representation of this fact on a log-log scale results in a straight line with slope J . An empirical plot of $\log Q(d)$ versus $\log d$ computed at various starting points t is called a pox diagram and appears as a uniform band of points centered about a straight line with positive slope J . The value $J = 1/2$ corresponds to a sequence of independent random variables. The values $J \neq 1/2$ represent various manifestations of long-term dependence. Before we can discuss this dependence in any more detail, it will be necessary to introduce some general classes of processes.

5. PROCESSES EXHIBITING LONG-TERM DEPENDENCE

A class which provides a natural framework within which to study long-term dependence is that of self-similar processes. A stochastic process $Z(t)$, $-\infty < t < \infty$ is self-similar with scaling exponent H if for all $a > 0$ $Z(at)$ and $a^H Z(t)$ have identical finite dimensional distributions. We will use the symbol \cong for

equality of finite dimensional distributions, as in

$Z(at) \cong a^H Z(t)$. The scaling exponent H is related to the level of long-term dependence. Let $X(i) = Z(i+1) - Z(i)$ denote the increments process derived from $Z(t)$.

If $Z(t)$ is a Brownian motion process, then the $X(i)$ are normally distributed and independent. Brownian motion is the special case of self-similar processes when $H = 1/2$. Therefore, this $Z(t)$ has $H = 1/2$ and its increments process $X(i)$ has $J = 1/2$. However, the R/S exponent for $Z(t)$ is $J = 1$.

A more general example of self-similarity is the class of stable processes developed by P. Levy with index $0 < \alpha \leq 2$. If $Z(t)$ is a stable process with $\alpha \neq 2$ then it has independent increments, infinite variance, and is self-similar with $H = 1/\alpha$. The increments process $X(i)$ has $J = 1/2$ whatever the value of H . For $\alpha = 2$ the process reduces to Brownian motion.

Fractional Brownian motion [Kolmogorov (1940), Levy (1953), Mandelbrot and Van Ness (1968)] is yet another example of a self-similar process. Recall that, for a particular integral type, fractional integrals are extensions to a continuum of the usual integer order. Ordinary Brownian motion is a certain integral (of order one) of white noise. Fractional Brownian motion is, therefore, a stochastic process that arises as a fractional integral of white noise. Let b_0 be a real number, let $H \in (0,1)$, and

let $B(t)$ denote ordinary Brownian motion. Define fractional Brownian motion $B_H(t)$, $t > 0$ as follows: $B_H(0) = b_0$

$$B_H(t) - B_H(0) = 1/\Gamma(H+1/2) \left[\int_{-\infty}^0 [(t-s)^{H-1/2} - (-s)^{H-1/2}] dB(s) \right. \\ \left. + \int_0^t (t-s)^{H-1/2} dB(s) \right].$$

Define $B_H(t)$ for $t < 0$ similarly. Kolmogorov (1940) and Levy (1953) are two early papers on such integrals. We can also view fractional Brownian motion as a realizable moving average, with weighting function $(t-s)^{H-1/2}$, on white noise.

For $Z(t)$, a fractional Brownian motion with parameter H , the increments process $X(i)$ is self-similar with self-similarity exponent H . If $Z(t)$ is (ordinary) Brownian motion then

$Z(t+s) - Z(t)$ has mean zero and standard deviation $s^{1/2}$ (i.e., an " $s^{1/2}$ law"). Almost all sample paths of fractional Brownian

motion are continuous and are not differentiable in the mean square sense for $H \in (0,1)$. The standard deviation of the increments (with Lag s) of $Z(t)$ with parameter H is $s^{2H} V(H)$ (an " s^H law") where

$$V(H) = [\Gamma(H+1/2)]^{-4} \left[\int_{-\infty}^0 [(1-s)^{H-1/2} - (-s)^{H-1/2}]^2 ds + 1/2H \right]^{1/2}.$$

These increments are called fractional Gaussian noise.

We will now summarize some properties of these processes. If an arbitrary self-similar process has stationary increments, and is continuous (in mean square) then $0 \leq H < 1$. If an arbitrary Gaussian self-similar process is non-constant, has stationary increments and continuity (m.s.) then it is fractional Brownian motion. Consider the relationship of J to H for some special classes of processes.

(i) $J = H$: $X(t)$ stationary, $E[X^2] < \infty$, $X^2(t)$ ergodic

$X^*(t)$ is self-similar H

($J = H = 1/2$ is the special case of $X(t)$ i.i.d.)

(ii) $J = H' - 1/2H' + 1/2$, $0 \leq J \leq 1$: $X(t)$ in the domain of attraction of a self-similar H' process

$X^2(t)$ in the domain of attraction of a self-similar H' process.

(iii) $J = 1/2$, $H = 1/\alpha$: $X(t)$ i.i.d., $E[X^2] = \infty$, $X(t)$ in the domain of attraction of a stable $0 < \alpha < 2$ process.

(iv) $J = 1$: $X(t)$ any non-stationary process that becomes stationary when differentiated (or differenced) one or more times.

(v) $J = H$: $Z(t)$ fractional Brownian Motion

$H \in [0,1]$.

There are other examples of self-similar processes such as Hermite processes and Rosenblatt processes which we will not discuss.

We have seen that $J \neq 1/2$ is a measure of long run dependence. What about the self-similarity exponent H . Assume $Z(t)$; $-\infty < t < \infty$ is self-similar with exponent $H \in (0,1)$; that $Z(0) = 0$, $E[Z(t)] = 0$, and $E[Z^2(t)] < \infty$; and that the

increments $X(i)$ are stationary. Then the correlation function of the $X(i)$ is given by

$$r(k) = 1/2 [|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}].$$

Clearly, for $H = 1/2$ the increments are i.i.d. For $H \neq 1/2$ we have two cases identified by Mandelbrot.

- (i) Persistent long run dependence: This is the case where $1/2 < H < 1$. Then $r(0) + r(1) + \dots = \infty$ and the spectral density is infinite at the origin.
- (ii) Antipersistent long run dependence: This is the case where $0 < H < 1/2$. Then $\dots + |r(-1)| + |r(0)| + |r(1)| + \dots < \infty$, and $\dots + r(-1) + r(0) + r(1) + \dots = 0$. The spectral density is zero at the origin.

6. STRONG DEPENDENCE, WEAK DEPENDENCE, AND MIXING

Previous sections have dealt with seemingly unrelated topics. These included asymptotic independence conditions, the rescaled adjusted range statistic, self-similar processes, and the definition of strong and weak dependence. In this section, it will be seen that these topics are more intimately related and have applications in theoretical physics.

A significant area of activity in theoretical physics has been the development of algebraic, in particular, group theoretic representations of fundamental physical systems. An important algebraic structure occurring in this work is the renormalization group [Benettin *et al.*, (1977)]. The motivation has been the modeling of critical phenomena, turbulence, and geophysical phenomena. The relationship to stochastic processes and dependence concepts arises when critical behavior and the renormalization group is interpreted in a probabilistic context. Cassandro and Jona-Lasinio (1978) develop this relationship. They define a process to be weakly dependent if it is g-mixing. Therefore, all of the "nice" processes that are ergodic, for which central limit theory holds, etc., are weakly dependent. Strong dependence is characterized by any process that violates g-mixing and corresponds to long-term dependence. Not only does long-term dependence and mixing have meaning in terms of physical phenomena but the dichotomy of persistent and antipersistent long run dependence has interpretations in terms of physical systems [Jona-Lasinio (1977)].

The physical interpretation of long run dependence leads to some interesting paradoxes in the interpretation of their spectral representations. In statistical language, the spectrum is a function which represents the distribution of the total variance of a

stochastic process over frequency. The frequency being the Fourier frequency of the orthogonal sinusoidal components. In physics, the variance is interpreted as energy and intervals in the spectral domain correspond to energy in the corresponding passband. Integrals, over their domain, of spectra of weakly dependent processes are finite. For strongly dependent processes, however, the spectra are not so nice [Taqqu (1980)].

The spectral density for this class of processes has a power law form which we will write as

$$s(\lambda) = |\lambda|^\alpha,$$

where λ is the frequency and exponent α is related to the underlying process. The nature of this relationship is not central and will be left to the references. Drawing an analogy to visible light, large $|\lambda|$ is considered in the ultraviolet region and small $|\lambda|$ is considered in the infra-red region. When $\alpha > -1$ the spectrum has heavy (non-integrable) tails and is said to exhibit the ultraviolet catastrophe. For $-1 < \alpha < 0$ the spectral density is infinite at the origin and exhibits the infra-red crisis. For $\alpha \leq -1$ the spectrum is non-integrable near the origin and this phenomena is called the infra-red catastrophe. When $\alpha = -1$ the process is called $1/f$ noise and has many physical interpretations [Voss (1979), Mandelbrot (1967)]. It is worth noting that many real life processes have been observed to have a power law spectra over extremely large frequency bands.

7. CONCLUSION

This discussion is only an introduction to the topic and is far from complete. For example, other measures of long run dependence have been developed with very different motivations. The Allan Variance of Time-and-Frequency metrology is mathematically different from the R/S statistic and was developed completely independently. The end result, its interpretation, and its graphical representation, however, is virtually identical. The relationships of the Allan Variance to the R/S statistic and of long run dependence to multivariate dependence concepts have also been considered.

REFERENCES

- Ahmed, A. N., Langberg, N. A., Leon, R. V. and Proschan, F. (1978). Two concepts of positive dependence with applications in multivariate analysis. Statistics Report M486, Department of Statistics, Florida State University.

- Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart, and Winston, New York.
- Benettin, G., DiCastro, C., Jona-Lasinio, G., Peliti, L., Stella, A. L. (1977). On the equivalence of different renormalization groups. In *New Developments In Quantum Field Theory and Statistical Mechanics*, M. Levy and P. M. Mitter, eds. Plenum, New York.
- Block, H. W., Savits, T. H. and Shaked, M. (1980). Some concepts of negative dependence. Technical Report.
- Blum, J. R. and Rosenblatt, M. (1956). A class of stationary processes and a central limit theorem. *Proceedings of the National Academy of Science, USA*, 42, 412-413.
- Cassandro, M. and Jona-Lasinio, G. (1978). Critical point behavior and probability theory. *Advances in Physics*, 27, 913-941.
- Esary, J. D. and Proschan, F. (1972). Relationships among some concepts of bivariate dependence. *The Annals of Mathematical Statistics*, 43, 651-655.
- Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, 38, 1466-1474.
- Hopf, E. (1937). *Ergodentheorie*. J. Springer, Berlin.
- Jona-Lasinio, G. (1977). Probabilistic approach to critical behavior. In *New Developments in Quantum Field Theory and Statistical Mechanics*, M. Levy and P. Mitter, eds. Plenum, New York. 419-446.
- Kac, M. (1959). *Statistical Independence in Probability, Analysis and Number Theory*. Carus Mathematical Monographs, 12, Wiley, New York.
- Kimeldorf, G. and Sampson, A. R. (1978). Monotone dependence. *Annals of Statistics*, 6, 895-903.
- Kolmogorov, A. N. (1940). Weinersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. Doklady, Academy of Sciences, U.S.S.R.*, 26, 115-118.
- Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and its Applications* 5, 204-208.
- Lawrance, A. J. and Kottegoda, N. T. (1977). Stochastic modeling of riverflow time series. *Journal of the Royal Statistical Society, A*, 140, 1-47.
- Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37, 1137-1153.
- Levy, P. (1953). Random functions: general theory with special reference to Laplacian random functions. *University California Publications in Statistics*, 1, 331-390.

- Mandelbrot, B. B. (1967). Some noises with $1/f$ spectrum, a bridge between direct current and white noise. *IEEE Transactions on Information Theory*, IT-13, 289-298.
- Mandelbrot, B. B. (1977). *Fractals - Form, Chance, and Dimension*. W. H. Freeman, San Francisco.
- Mandelbrot, B. B. and Taqqu, M. S. (1979). Robust R/S analysis of long run serial correlation. IBM Technical Report RC7936, No. 34217.
- Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motion, fractional noises and applications. *SIAM Review*, 10, 422-437.
- Mandelbrot, B. B. and Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5, 967-988.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences, USA*, 42, 43-47.
- Rosenblatt, M. (1961). Independence and Dependence. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability, II*, J. Neyman, ed. University of California Press, Berkeley and Los Angeles. 431-443.
- Ross, S. M. (1979). Multivalued state component systems. *Annals of Probability*, 10, 379-383.
- Taqqu, M. S. (1980). Self-similar processes and related ultra-violet and infrared catastrophes. To appear in *International Colloquium on Random Fields*, Esztergom, Hungary, June 24-30. Janos Bolyai Mathematical Society and North Holland Publishing Co.
- Voss, R. F. (1979). $1/f$ (flicker) noise: A brief review. *Proceedings of the Thirty-Second Annual Frequency Control Symposium*, 40-46.
- Zhurbenko, I. (1975). Mixing conditions and asymptotic theory of stationary time series. In *Stochastic Processes and Related Topics*, M. L. Puri, ed. Academic Press, New York. 259-265.

[Received January 1981]

SOME DISTRIBUTION THEORY RELATED TO THE ANALYSIS OF SUBJECTIVE PERFORMANCE IN INFERENTIAL TASKS

J. AITCHISON

Department of Statistics
University of Hong Kong

SUMMARY. When inferences made by a subject differ from normative statistical inferences it is possible to define measures of this divergence which throw considerable light on the nature of the deficiencies in the subjective performance. These measures involve quantifying differences between distributions. For the analysis of a single subject performing a single task the paper generalizes previous work by replacing absolute measures by more meaningful relative measures and by extending the scope of the inferential tasks involved. For a single subject performing many tasks and for groups of different subjects statistical analysis requires a rich parametric class of distributions to describe patterns of variability of probabilistic data, and the use of logistic-normal distributions is advocated for this purpose. Simple illustrations are provided of the main analytical techniques.

KEY WORDS. Degree of uncertainty, feature selection discrepancy, inference discrepancy, inferential tasks, information gain index, logistic normal distributions, normative models, performance analysis, probabilistic data, subjective inference.

1. INTRODUCTION

Many individuals in the course of their routine work have to make important intuitive or subjective inferences on the basis of similar information in repetitive but independent situations. For example, a clinician on the basis of his past experience and on the information provided by symptoms, signs and diagnostic tests has to judge the relative plausibilities of possible diseases for each patient. A personnel manager, on the basis of a curricu-

lum vitae and information and impressions gained at interview, has to assess the chances of success in the post advertized for each applicant. A process controller having some picture of the variability of a characteristic of a manufactured item relative to the quality of the material input, has to infer a likely range for the characteristic of a new item with known material input. A biochemist on the basis of the circles cleared in an infected medium by droplets of an antibiotic of known standard concentrations and the circles cleared by droplets of a blood specimen of unknown concentration makes inferences concerning this unknown concentration in an often subjective way from graphical considerations.

For some of these inference tasks the person concerned may have recourse to statistical help or to a computer package but this is not invariably so. Moreover it is of intrinsic interest to study the ability and variability of individuals in various professions in making appropriate inferences subjectively. It is the purpose of this paper to describe the underlying concepts of distribution theory and useful methods of analyzing the performance of such subjects in inferential tasks related to their work.

Apart from any psychological, inter-disciplinary or cross-cultural insights such performance analyses may provide we have found them an excellent vehicle for statistical education. The author has now some ten years' experience of presenting inferential tasks to students of statistics at various stages of study, in service courses to students of other disciplines, to school children from fourth to sixth forms, to civil servants, clinicians, physicists, historians and even to statistician colleagues. In his admittedly subjective judgement the immediacy of the challenge of such an inferential task, particularly when presented in a familiar context, stimulates in the subject a keen interest. Moreover subjects, so often finding their own performance far from adequate and at variance with their peers, appreciate that statistical methodology may have much to give to their particular discipline and look forward to learning the 'secrets' of proper inferential methods.

2. INFERENCE TASKS, STATEMENTS AND TRIALS

We first establish a terminology, notation and framework for the study of subjective performance analysis in inferential tasks.

Inferential task. In an inferential task (such as medical diagnosis or antibiotic assay) a *subject* (clinician, biochemist) is presented with a *case* (patient, blood-sample) for which an *inferential statement* (diagnostic assessment, antibiotic assay) con-

cerning the true hypothesis or *index* (true disease type, true concentration of antibiotic) is required. The subject is aware that the case has associated with it a unique but unknown index belonging to a known index set T (set of feasible diseases assumed or defined to be mutually exclusive, range of possible antibiotic concentrations). To help him arrive at his inferential statement for a particular case the subject has available information concerning the case in the form of data on a number of *features* (results of diagnostic tests, clearance circle diameters) which can thus be regarded as a feature vector in some defined feature vector space X .

Inferential statement. An inferential statement assigns a weight $w(t)$ to each index $t \in T$ in such a way that $w(t_1)/w(t_2)$ can be interpreted as the relative probability or plausibility of t_1 compared with t_2 . We can usually arrange for the weighting function to correspond to a probability or plausibility function $p(t)$ over T , for example, by considering $p(t) = w(t)/\sum_T w(t)$ or $w(t)/\int_T w(t)dt$.

Previous experience and training sets. The subject may already have some training and experience in the kind of inferential task under study, but this is seldom quantifiable. Examples again are clinician and biochemist with skills in diagnosis and assay. But inferential tasks can be selected so that the relevant experience and training are under the control of the experimenter, and hence quantifiable. For example, where diagnostic tests unfamiliar to the clinician have been evolved he can be presented with information on the complete *training set* of cases whose diagnoses and test results are known. In the antibiotic assay a training set is an essential ingredient of the task; since clearance diameter is known to vary from batch to batch of infected medium, it is essential for the subject to know the concentrations and clearance diameters of a training set of cases, referred to in assay work as 'the standards'.

We write D to denote the data of a training set of m cases for each of which the true index and feature vector are known. Then D takes the form

$$D = \{(t_i, x_i) : i = 1, \dots, m\},$$

where $t_i \in T$, $x_i \in X$ ($i = 1, \dots, m$).

Inferential trials. In an inferential trial a subject is presented with a *test set* of n unrelated or independent cases and on the basis of their feature vectors y_1, \dots, y_n is asked to make inferential statements about their unknown indices u_1, \dots, u_n . We

shall concentrate mainly on trials where the inferential statements required of the subject are composite in nature, say density functions s_1, \dots, s_n on T . The performance data thus consist of the set

$$\{(y_i, s_i) : i = 1, \dots, n\}.$$

Having defined our terms we can now briefly describe previous work in the analysis of subjective performance in inferential tasks. Note that we confine the term inferential task to the provision by the subject of the equivalent of a probability distribution over the complete set of possible indices or hypotheses. In particular, we are not concerned with subjective studies involving other forms of inference, such as assigning scores on a psychotic scale to profiles of psychiatric patients (Goldberg, 1970), or of decision-making, such as in artificial tasks (Davidson *et al.*, 1957) or in medical treatment decisions (Aitchison *et al.*, 1973). Moreover our emphasis is on the analysis of actual performance and not on devices to encourage good probability estimation (Good, 1952), though measures of performance and penalties for bad estimation are, of course, related.

Early studies of performance in inferential tasks as defined here are by Edwards and Phillips (1964) and Phillips and Edwards (1966), who report conservative use of information in simple inferences concerning the composition of bags containing colored poker chips. For an extreme case of such conservatism among American lawyers, see Raiffa (1968, p. 20). Taylor *et al.* (1971) move from the artificial context of chips in bags to real problems of medical diagnosis and try to face clinicians with realistic tasks of inferring diagnoses from sequentially acquired information on patients. Taking a statistical diagnostic system as a norm they then analyse performance in terms of various measures of discrepancy from normative assessments, one of these measures being a generalization of the Edwards-Phillips simple version of conservatism. Technical definitions and discussion of these measures and their use in other inferential tasks are to be found in Aitchison (1974), Aitchison and Kay (1973, 1975) and Kay (1976).

3. MEASURES OF NORMATIVE COMPARISON FOR A SINGLE SUBJECT PERFORMING A SINGLE TASK

Normative model and system. When it is possible to formulate a rational model according to which a subject with the given information ought to be making his inferences then we can compare the subject's actual conclusions with the corresponding normative inference. In situations where such a normative model can be specified then standard procedures can be applied to use the data D of the training set to obtain a fitted model or *normative system*

which may be applied to the cases of a test set to produce inferential statements about these cases. The technical statistical details of the construction of a normative system need not concern us until we consider specific areas of study. A normative system can thus be expressed in the form of a conditional density function

$$p(u|y,D)$$

over T , for a test case with feature vector y .

When the normative system is applied to the feature vectors y_1, \dots, y_n of the test set it produces normative statements, say r_1, \dots, r_n .

Nature of comparisons. Comparison of how a subject's performance departs from normative performance thus requires the construction of measures of the extent of the departure of statements s_i from corresponding normative statements r_i ($i = 1, \dots, n$).

In the tasks so far defined we have considered the feature vector information being supplied in one piece and there being a single final inferential statement. Later we shall consider tasks where the feature vector information is supplied in sequence to the subject and he is asked to make an inference statement after each step in the sequence. In such circumstances when considering a typical step we shall have information about the relative probabilities assigned to the unknown index before as well as after the feature information given to him. To deal with this at the present stage of our discussion we therefore suppose that for each case i the subject makes an inferential statement q_i prior to receiving the feature vector y_i for the case ($i = 1, \dots, n$).

For a single test case or inferential task therefore we have to suppose that there is a prior inferential statement $q(t)$ on T and that on the basis of knowledge of the feature vector y for the case we have to compare the subsequent subjective inferential statement $s(t)$ on T with the corresponding normative inferential statement $r(t)$ on T .

Later, in Sections 6 and 7, we shall be concerned with variability in $s(t)$ from task to task within the same subject and between subjects performing the same task. For the moment we confine attention to meaningful summary characteristics or measures of performance, along the lines of those introduced by Taylor *et al.* (1971). Their characteristics are, however, measured on an absolute scale, not related to what is directly achievable within the inferential task, and so often difficult to interpret. A first purpose of this paper is to replace them where necessary by measures defined on a more meaningful relative scale.

Degree of uncertainty. Associated with any composite inferential statement, say $q(t)$ on T , there is the now familiar information theory concept (Shannon, 1948; Khinchin, 1957) of a degree of uncertainty $U\{q(t)\}$ or $U(t)$ remaining in the identification of the true index:

$$U(q) = -\sum_T q(t) \log q(t) \text{ or } -\int_T q(t) \log q(t) dt \quad (1)$$

where $q(t) \log q(t) = 0$ if $q(t) = 0$.

Inference discrepancy. Since the subject records an inferential statement $s(t)$ on T which, according to the normative system, should be $r(t)$ on T we require, in order to assess his ability in inference, an overall measure of the difference between $s(t)$ and the target $r(t)$. This is provided by an information theory measure $I(r, s)$ due to Kullback and Leibler (1951):

$$I(r, s) = \sum_T r(t) \log \frac{r(t)}{s(t)} \text{ or } \int_T r(t) \log \frac{r(t)}{s(t)} dt, \quad (2)$$

with the property that

$$\begin{aligned} I(r, s) &> 0 \quad (s \neq r) \\ &= 0 \quad (s = r). \end{aligned}$$

Information gain index. We can quantify such notions as 'under-using the information available', 'reading too much into the data', 'going contrary to the evidence' in terms of an information gain index $G(q, r, s)$. Suppose that $U(q) > U(r)$ so that the normative system has removed $U(q) - U(r)$ of uncertainty or equivalently gained this amount of information about the index. The subject on the other hand has gained an amount $U(q) - U(s)$ of information in his move from q to s . Consider now the ratio

$$G(q, r, s) = \frac{U(q) - U(s)}{U(q) - U(r)}. \quad (3)$$

If $G(q, r, s) > 1$ then the subject has removed more uncertainty than the normative move and so can be said to be acting liberally or reading too much into the data. If $0 < G(q, r, s) < 1$ the subject is acting conservatively or underusing the data. If $G(q, r, s) < 0$ then the subject is increasing the uncertainty when he ought to be decreasing it and we can say that he is running contrary to the evidence.

The same kind of argument applies to $G(q, r, s)$ when $U(q) - U(r) < 0$, and in the special circumstance when $U(q) - U(r) = 0$ we set $G(q, r, s) = +\infty$, 1 or $-\infty$ according to whether $U(q) - U(s)$ is positive, zero or negative.

Feature selection discrepancy. In a number of inferential tasks the subject may be faced not only with problems of updating an inferential statement on the basis of an observed feature vector y , but also that of selecting which feature from a set of alternatives he would like to observe. For example, in diagnosis the clinician would almost certainly have to choose which of a number of diagnostic tests should be carried out.

Suppose that from a starting density function $p(t)$ on T any one of a set F of features is available. Consider the choice $f \in F$. If outcome x is observed and leads to a normative posterior assessment $p(t|x)$ then the reduction in uncertainty or gain in information is $U\{p(t)\} - U\{p(t|x)\}$. In comparing the relative merits of different feature selections we do not know the outcome and so, following Lindley (1956), we have to measure the merit of f in terms of the expected gain of information for f from starting density function $p(t)$:

$$H\{f, p(t)\} = \sum_x [U\{p(t)\} - U\{p(t|x)\}] p(x). \quad (4)$$

The larger this is the more informative the feature is and so a normative choice $f^* \in F$ satisfies

$$H\{f^*, p(t)\} = \max_F H\{f, p(t)\}. \quad (5)$$

Note that f^* depends on $p(t)$: what is an optimum choice from one $p(t)$ may be poor from some other starting position. If a subject, at a declared assessment $p(t)$ and faced with a choice of experiment, chooses f then the amount by which the expected gain of information for f falls short of the expected gain of information for f^* gives the measure used by Taylor *et al.* (1971):

$$H\{f^*, p(t)\} - H\{f, p(t)\}.$$

We can, however, more appropriately compare this with the subject's worst possible choice f_* , where

$$H\{f_*, p(t)\} = \min_F H\{f, p(t)\},$$

to obtain a relative measure of feature selection discrepancy:

$$S\{f, p(t)\} = 1 - \frac{H\{f^*, p(t)\} - H\{f, p(t)\}}{H\{f^*, p(t)\} - H\{f_*, p(t)\}} \quad (6)$$

The measure S is confined to the range $0 \leq S \leq 1$, the value 1 corresponding to the worst possible selection and the value 0 to the normative selection.

Measures associated with normal assessments. For T a finite or discrete set, such as in the diagnostic inferential tasks already cited, the computations of the measures described above are comparatively simple summations. When, as in prognostic, assay and calibration studies, T is the real line or even a higher dimensional real space then evaluations of the measures for univariate and multivariate normal assessments and distributions can prove useful either in their own right or as approximations. Using the notation $N_d(\lambda, \Sigma)$ for a d -dimensional normal distribution with mean vector λ and covariance matrix Σ we have the following results.

$$U(q) = \begin{cases} \frac{1}{2}\{1 + \ln(2\pi\sigma^2)\} & \text{when } q(t) \text{ is } N_1(\lambda, \sigma^2) \\ \frac{1}{2}\{d + \ln \det(2\pi\Sigma)\} & \text{when } q(t) \text{ is } N_d(\lambda, \Sigma). \end{cases} \quad (7)$$

When $r(t)$ is $N_d(\lambda, \Sigma)$ and $s(t)$ is $N_d(\mu, \Omega)$ then

$$I(r, s) = \frac{1}{2}\{\text{trace}(\Omega^{-1}\Sigma) - \ln \det(\Omega^{-1}\Sigma) - d\} + \frac{1}{2}(\lambda - \mu)' \Omega^{-1}(\lambda - \mu). \quad (8)$$

The simplification for the univariate case when $r(t)$ is $N_1(\lambda, \sigma^2)$ and $s(t)$ is $N_1(\mu, \omega^2)$ is:

$$I(r, s) = \frac{1}{2}\left\{\left(\frac{\sigma}{\tau}\right)^2 - \ln\left(\frac{\sigma}{\tau}\right)^2 - 1\right\} + \frac{1}{2}\left(\frac{\lambda - \mu}{\omega}\right)^2. \quad (9)$$

Note that the first bracketed part separates out a component of the inference discrepancy which measures departure of the subject's assessment of the covariance structure from the normative value. The second component does not, however, give a pure measure of the disagreement in means because of its involvement with the subject's variance or covariance assessment ω^2 or Ω .

Since $G(q, r, s)$ is a simple construction of U values there is no need to provide an explicit expression.

When $p(t)$ is $N(\lambda, \Sigma)$ the form of $H\{f, p(t)\}$ depends on whether in the inferential task it is more appropriate to specify $p(x|t)$, say as $N(\alpha + Bt, \Gamma)$ or to specify $p(t|x)$ as $N(\lambda + \Delta x, \Omega)$. In the first case

$$H\{f, p(t)\} = \frac{1}{2}\ln \det(I + \Gamma^{-1}B\Sigma B') \quad (10)$$

and in the second case

$$H\{f, p(t)\} = \frac{1}{2}\ln \det(\Omega^{-1}\Sigma). \quad (11)$$

4. SEQUENTIAL INFERENCE TASKS

The measures of performance of Section 3 have been defined on the basis of the feature vector being presented as a whole for a single inference task. If it is meaningful to present the feature vector components one at a time or in successive blocks then the subject can be faced with a sequential inference task, being required to update his initial assessment $q(t)$ immediately after each component x_1, \dots, x_k has been presented to him resulting in successive subject assessments, say $s_1(t), \dots, s_k(t)$. We can then clearly analyze his performance after each such subjective assessment.

Such a sequential performance analysis can take two forms. The first is a relative one in which at the i th stage we treat the subject's present view $s_{i-1}(t)$ as the starting $q_i(t)$ in the evaluation of the normative assessment $r_i(t)$ against which $s_i(t)$ is to be judged. Secondly, there is an accumulative or absolute performance analysis which in the normative updating to obtain $r_i(t)$ after the i th stage uses as starting assessment the previous normative updating $r_{i-1}(t)$ rather than the subject's $q_i(t)$. Which is the more appropriate will depend to some extent on the nature of the particular inferential task. On the whole we prefer the relative analysis because it builds successively on the subject's immediately held belief and so has a greater opportunity of identifying particular circumstances in which discrepancies from the normative occur.

The definition of feature selection discrepancy lends itself easily to sequential inference tasks. At each stage of a sequential inference task, instead of presenting the subject with the next component, we may ask him to choose which of the components not so far revealed is likely to clarify the uncertainty most. At the i th stage, with his current subjective assessment at $s_i(t)$, if he chooses feature f_i from the set F_i of features available then replacement of f , F and $p(t)$ by f_i , F_i and $s_i(t)$ respectively in the definition of S at (6) produces the appropriate features selection discrepancy for the i th stage.

5. SOME SPECIFIC INFERENTIAL TASKS

In this section we refer to some inferential tasks already partially reported in the literature, describe briefly some others which we have tested in pilot studies, and indicate some other

applied areas which may be worthy of investigation. In particular, we shall draw attention to one or two interesting and unexpected aspects of the associated performance analysis which we consider to call for further more careful investigation.

Diagnostic inferential tasks. A number of such tasks have been reported by the author and colleagues so that there is no need to give more than the relevant references and a few brief comments. First, we emphasize that in these tasks diagnosis is presented as an inference rather than a decision problem, the subject being required, for a sequence of patients, to assign probabilities to the possible disease types on the basis of patient information released to him either sequentially or as a whole. For a valid performance analysis it is necessary to know exactly what information about a case is known to the subject. It is therefore not possible to allow the subject to see the patient lest he collects visual or other information unknown to the analyst and so information must be supplied verbally or possibly at some visual display unit. To the extent that there is no contact with the patient it could be claimed that such studies do not put clinician subjects into their natural inference-making setting but most subjects regard the tasks presented as fair tests of diagnostic skills. Moreover when interest is in comparing the inferential skills of clinicians with those of other professions direct access to patients is clearly not possible.

Performance analysis studies of diagnostic inference differ in a number of respects:

- (i) the extent to which the experience of the subject in the particular diagnostic area has already been acquired and so is not determinable or can be completely supplied by the analyst (see the previous comments in Section 2);
- (ii) the extent to which the information on a new case can be supplied sequentially;
- (iii) in a sequential task the extent to which the choice of the next feature is required of the subject;
- (iv) the extent to which any assumptions of the normative assessments are valid.

An early study in this area (Taylor *et al.*, 1971) was concerned with differential diagnosis of non-toxic goitre, depended on the self-attained experience of six clinicians in a specialist clinic, was sequential in nature with the clinician at each stage choosing the next feature. Each of the features was categorical, with either two or three categories, and for the normative model the $p(x|t)$ probabilities were assessed on the basis of data on about 50 cases for each of three types and on

the assumption that for given type features are independent. To the extent that no allowance was made for sampling error and that the independence assumption was undoubtedly suspect this study may be criticized. Nevertheless as an expository aid for inferential principles in diagnosis, particularly with its strong visual appeal in the presentation of sequential diagnostic inference as paths within a triangle or triangular bowl of uncertainty, it has proved invaluable. One of the awkward features in this example is undoubtedly the difficulty in obtaining simple models for multivariate categorical data, though with greater familiarity of, and availability of programs for, log-linear analyses and kernel density function estimation, more appropriate normative models could be devised.

This difficulty of modelling the dependence of the features does not arise when the features are quantitative and when the feature vectors, or possibly transformed feature vectors, for given type, follow a multivariate normal form, since for example in sequential inference conditional dependence can be easily expressed and programmed. Details of these technical aspects may be obtained in Aitchison and Kay (1975) and Kay (1976). Aitchison and Kay (1973) presented a non-sequential differential diagnostic problem as a competition. 'Experience' took the form of details of the six-dimensional features of twelve cases of each of three types, this training set and the subsequent cases being generated by multivariate normal simulations. Another useful instructive inferential task is another simulative situation variously presented as Doctor's Trilemma or differential diagnosis of three forms of Newmath Syndrome. In this 'experience' consists of informing the subjects in simple terms that the ten features are, for given type, independent binary and each subject has the thirty binary probabilities in front of him as he updates in the diagnostic triangle as the binary features of the case are revealed to him sequentially. For further details, see Aitchison and Kay (1973) and Aitchison (1974).

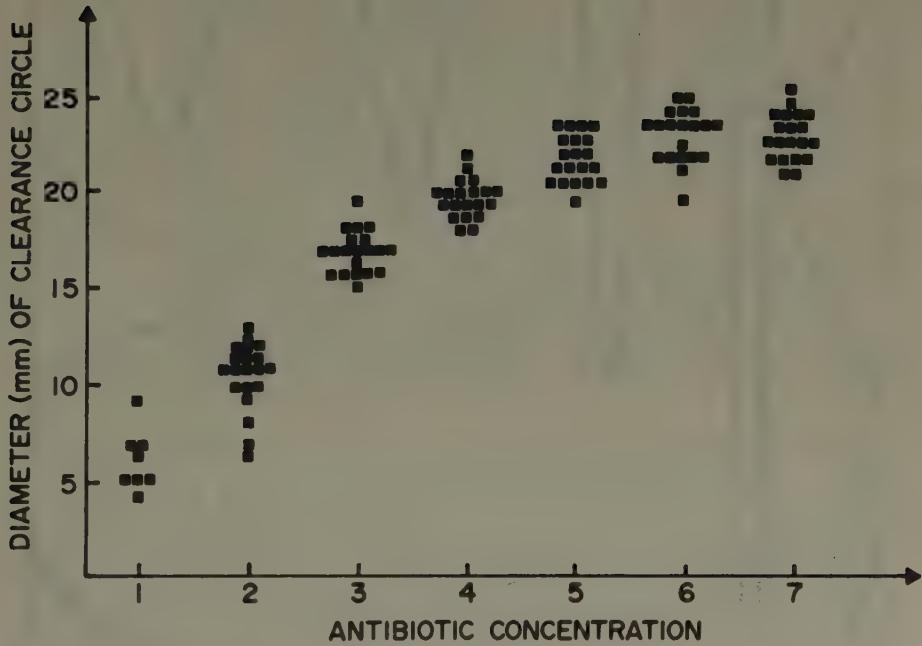
It is clear that any real or simulated task in this diagnostic area can be easily presented, the only constraint being that for real diagnostic tasks we have available an appropriate and sufficient training set on which to base normative assessments. One aspect of normative assessments worth noting is that predictive rather than estimative assessments in the sense defined by Aitchison *et al.* (1977) are to be recommended.

Predictive and prognostic inferential tasks. Since we shall be describing below a calibrative inferential task which calls for a density-function type assessment similar to those required here we shall confine ourselves to the bare outline of such tasks.

Suppose that we give a subject who is aware of the concept of the normal density function a set of observations from a normal distribution set out on a horizontal scale and then pose him the following task. Another observation is about to be recorded. Can the subject draw a pattern of plausibilities, essentially an unscaled density function, on the horizontal scale given, whose heights show his subjective assessments of the relative plausibilities of the various possible values? This requires a composite assessment which we can convert to proper density function form $s(t)$ to be compared against a normative assessment $r(t)$. One point worth noting here is that the usually fitted normal curve with sample mean and standard deviation substituted is not an appropriate normative assessment and is better replaced by a Student-type density function for the reasons given in Aitchison (1975).

More complicated inferential tasks here involve regression-type situations. In such a task the set T will usually be the real line, the set of possible response or dependent variables, while X is one- or higher-dimensional and consisting of possible explanatory, concomitant or independent variables. A typical simple inferential task in this area is to provide the subject with a regression-type scatter diagram with x -axis horizontal and t -axis vertical, then to ask the subject, after suitable explanation of the meaning of the task, to provide a density function $s(t)$ or 'pattern of plausibility' for the possible t values corresponding to a given x .

Calibrative inferential tasks. The type of task here is best described in terms of a specific simple example which we have given to a variety of subjects with some very interesting results. For this task each subject receives a copy of Figure 1 which is the training set, data for the 'standard curve' for an assay or calibration, and the background to the problem is explained to the subjects. The problem concerns the assay of the concentration t of an antibiotic in a patient's blood. Droplets of standard preparations of known concentrations t_i of the antibiotic are placed on a prepared infected medium on Petri dishes and, after cooking for 24 hours, the diameters x_i (mm) of the circles cleared, which are of course related to the concentration but in a statistical rather than a deterministic way, are recorded. In Figure 1 these (t_i, x_i) points are plotted. The subject is then made aware that the problem is to try and infer something about the unknown concentration of antibiotic in the patient's blood from knowledge only of the diameter of the clearance circle from a single droplet. He is invited to make use of the numbered patterns of variability shown in Figure 2 and supplied to him on a transparent sheet,



TASK NO.

1. MOST PLAUSIBLE STIMULUS

CURVE NO.

2. MOST PLAUSIBLE STIMULUS

CURVE NO.

FIG. 1: Scattergram of (antibiotic concentration, clearance diameters) for standards in calibration task.

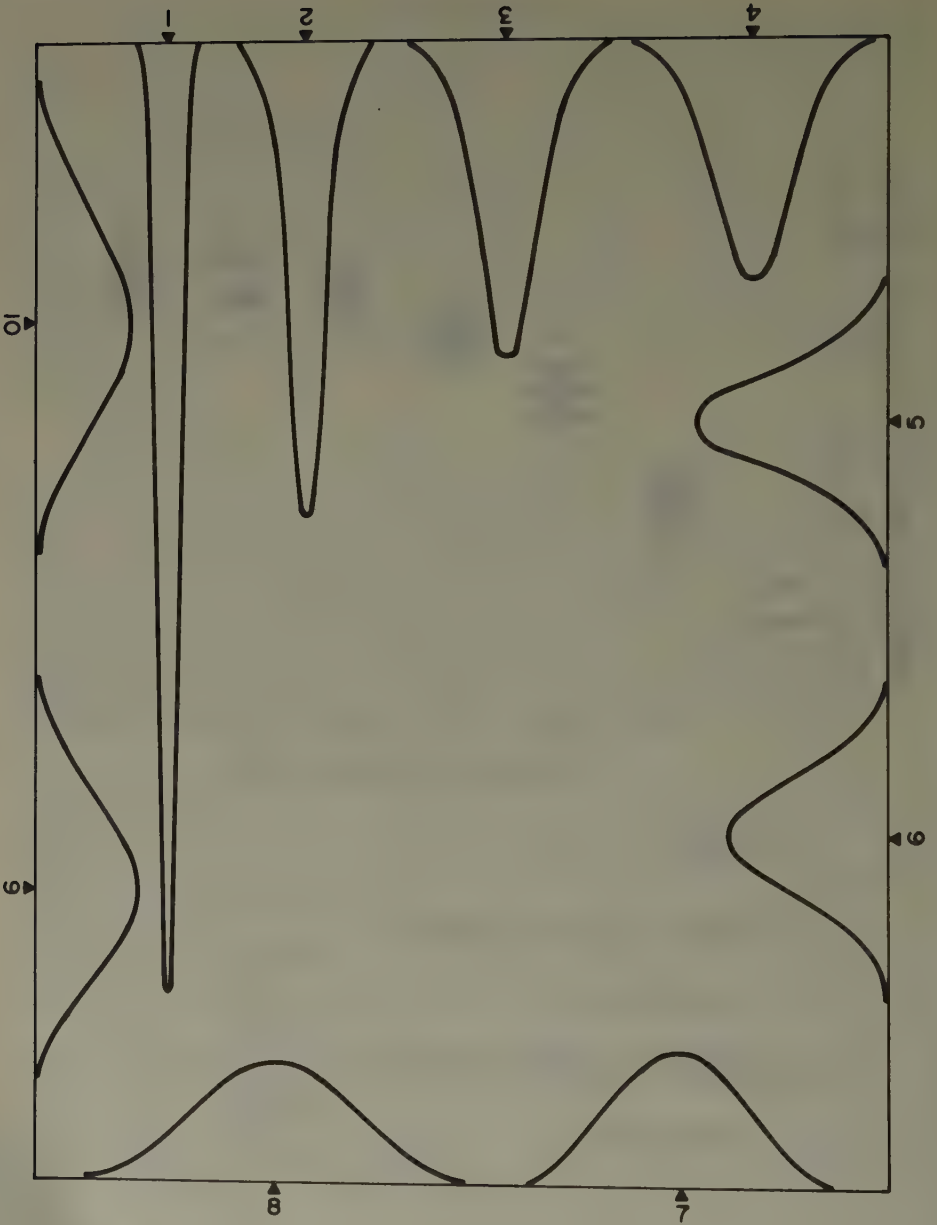


FIG. 2: Patterns of variability for subjective inference in calibration task.

to place what he regards as an appropriate pattern on the horizontal t-axis. The meaning of such patterns is explained to him in some detail, for example,

- (i) that the mode of the pattern selected should naturally be placed above the concentration regarded as most plausible;
- (ii) that with such patterns the relative heights of the curve above any two concentrations should reflect the subject's view of the relative plausibilities of these concentrations;
- (iii) that the 'narrower' the pattern chosen the more precise the subject is regarding the method of assay;
- (iv) that, since only a finite number of patterns can be provided, he is free to choose an in-between pattern by 'interpolation', recording his choice to one decimal place; for example, pattern 4.3 is intermediate between pattern 4 and pattern 5 but nearer to 4 than 5.

At the outset the subject is told that all concentrations are equally likely. In our studies two tasks are given. First the subject is told that the diameter from a single droplet of the patient's blood was 19 mm. and is asked to identify his pattern by writing down the most plausible value and the pattern curve number. In the second he is told three diameters of 18.5, 18.5 and 20 mm. Again he is asked to select, on the basis of this information, his pattern by again noting his chosen most plausible value and pattern curve number. The assessments of performance can, of course, be easily quantified by comparison of the selected curve $s(t)$ with a normative curve $r(t)$, such as the calibrative density function of Aitchison and Dunsmore (1976, Chapter 10) or a normal approximation to it. The measures (8)-(11) are then appropriate.

One interesting and surprising feature of the results is that in each of a number of different groups - statistical students in various years, clinicians, physicists - approximately one half choose a wider pattern in the second task than in the first, contrary to the common sense view that more experimentation should provide a more precise inferential statement. It is clearly a phenomenon which is well worth further investigation. One possible explanation is that with the single diameter some subjects have a tendency to forget about or underestimate the variability in diameter for given dose, whereas when they are presented with three diameters showing variability they take account of this variability.

Further more thorough and more extensive inferential task studies are being prepared particularly in this area of calibration. For example, it will be interesting to discover how skillful subjects are in choosing between different response variables - for example, do they appreciate the role played by the slope of the response line and the variability about the line? It is even possible to present inferential tasks involving both calibrative and diagnostic skills.

6. DISTRIBUTIONS OF INFERENTIAL STATEMENTS

When studying a single subject performing differential tasks or a number of different subjects performing the same task we will be faced with a set of inferential statements s_1, \dots, s_n , each a probability distribution over the set T .

For the statistical analysis of such data it is clearly an advantage to consider distributions of probability distributions over T . Consider first the situation when T is finite $(d+1)$ -dimensional. Then s_1, \dots, s_n are probabilistic data in the sense that they can be represented by points in the d -dimensional simplex

$$S^d = \{u : u_i > 0 \quad (i = 1, \dots, d), \quad u_1 + \dots + u_d < 1\}.$$

Figure 3 shows the inferential statements of 56 statisticians, undertaking the same inferential task in a diagnostic competition at a conference in multivariate analysis; Figure 4 shows the corresponding statements of 11 second-year undergraduates.

Distributions over the simplex have been recently discussed by Aitchison and Shen (1980) and Aitchison (1981), who advocate the class of logistic-normal distributions as a richer, more statistically tractable class than the more familiar Dirichlet class. If $u \in S^d$ then $v_i = \ln(u_i/u_{d+1})$ ($i = 1, \dots, d$), where $u_{d+1} = 1 - u_1 - \dots - u_d$, defines a vector v in R^d . If v follows a $N_d(\mu, \Sigma)$ distribution then u is said to follow a corresponding logistic-normal distribution $L_d(\mu, \Sigma)$ over S^d . The term logistic-normal is used since u is related to v by the logistic transformation

$$u_i = e^{v_i} / (1 + e^{v_1} + \dots + e^{v_d}) \quad (i = 1, \dots, d).$$

There are, indeed, some grounds for expecting that the pattern of variability of inferential statements may follow logistic-normal distributions. If, for given t , the distribution of the components x_1, \dots, x_k of the feature vector are

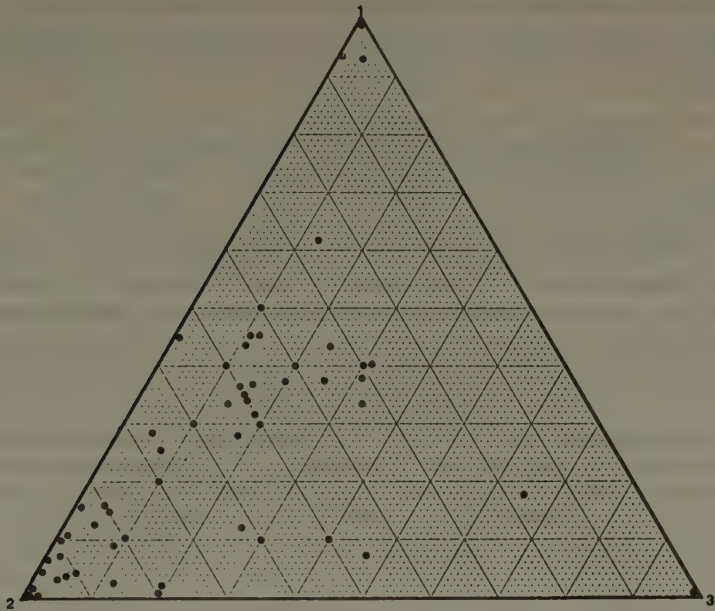


FIG. 3: Diagnostic inferences of 56 statisticians.

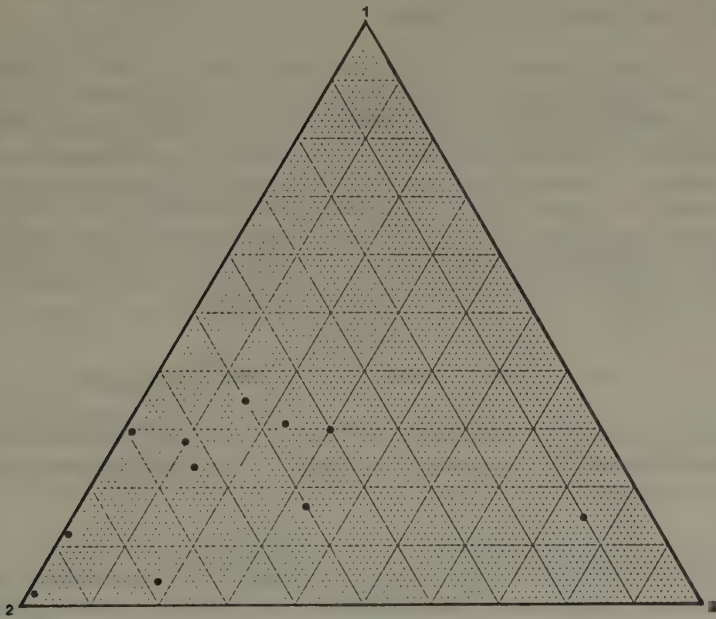


FIG. 4: Diagnostic inferences of 11 second-year undergraduates.

independently distributed with density functions $p_j(x_j|t)$ then from prior probabilities $p(t)$ ($t \in T$) we have, by Bayes's formula,

$$v_i = \ln \frac{p(t_i|x)}{p(t_{d+1}|x)} = \ln \frac{p(t_i)}{p(t_{d+1})} + \sum_{j=1}^k \ln \frac{p_j(x_j|t_i)}{p_j(x_j|t_{d+1})}.$$

Provided that the distributions $p_j(x_j|t)$ satisfy sufficient regularity conditions for the central limit theorem to be valid and that k is a reasonable size, then the v_i , being sums, will tend to be jointly normally distributed, and so the true inferential statements will follow logistic-normal variability. Of course in practice the independence assumption will not be valid, so that logistic-normality of true inferential statements would have to rely on a central limit property for dependent sums. But we suspect that subjects have considerable difficulty in taking account of dependence in their assessments so that if their subjective process does correspond to some rough and ready form of Bayes's formula it is likely to be in an approximately independent form. All this is speculative and it seems doubtful whether the subjective process can ever be investigated in this amount of detail. But at least there is a *prima facie* case for investigating logistic-normality of distributions of inferential statements of finite T .

For T non-finite, such as the real line in the calibration problem, the description of the variability of inferential statements, now probability distributions over the real line, is much more difficult. If the task takes the form of the selection of a normal curve then an inferential statement is equivalent to (m, s) , where m and s are the mean and standard deviation of the selected normal curve. For this task we then have to select some suitable joint distributions for (m, s) and it is possible that some normal-Wishart form may be appropriate.

7. TWO STUDIES INVOLVING INFERENTIAL TASKS

We now illustrate the application of logistic-normal analysis to the inferential statements of groups of subjects, performing diagnostic tasks.

7.1 Doctor's Trilemma. The subjects were 48 first-course statistics students and each was presented with four tasks, the four cases requiring a diagnosis between three types with information from the ten binomial tests presented sequentially. For details of the training set and cases presented see Aitchison (1974).

Each of the 24 possible orders of presentation of the cases was allocated to two students, the allocation being at random. The trial was conducted in two sessions. At the first session, early in the course and before students had met the appropriate technical tool of Bayes's formula each student tackled his first two cases. The remaining two cases were presented at the second session, some six weeks later and after meeting Bayes's formula in lectures. Subjects were not informed that Bayes's formula was the appropriate tool and were not allowed to write anything on paper except their inferential statements.

Kay (1976) has analyzed the various measures of performance of Section 3 separately for this study. Here the analysis applies to the complete or inferential statement and not just single aspects. If, s_{ij} denotes the inferential statement of the i th subject on the j th task presented to him then a model Ω for the analysis of the various possible effects is as follows:

$$s_{ij} \text{ is } L_2(\mu + \alpha_i + \beta_j + \gamma_{K(i,j)}, \Sigma),$$

where $K(i,j) = 1$ or 2 according as the case comes before or after knowledge of Bayes's formula. The usual form of identifiability restrictions apply:

$$\sum_{i=1}^{48} \alpha_i = 0, \quad \sum_{j=1}^4 \beta_j = 0, \quad \gamma_1 + \gamma_2 = 0.$$

Here the α_i and β_j denote subject and task effects and non-zero values of γ_1 and γ_2 will indicate some effect associated with knowledge of Bayes's formula.

For testing any hypothesis ω within the model Ω the usual chi-squared approximation at significance level α form of the generalized likelihood ratio test can be expressed in the form

$$192 \log \frac{\det \hat{\Sigma}_{\omega}}{\det \hat{\Sigma}_{\Omega}} > \chi^2(r, \alpha),$$

where $\hat{\Sigma}_{\omega}$ and $\hat{\Sigma}_{\Omega}$ are the maximum likelihood estimates of Σ under ω and Ω , r is the number of independent constraints on the parameters required to specialize Ω to ω and $\chi^2(r, \alpha)$ is the upper α point of the chi-squared distribution with r degrees of freedom. Figure 5 gives the complete lattice of hypotheses with the test quantities and their critical values. Moving down the lattice we can reject all the hypotheses except

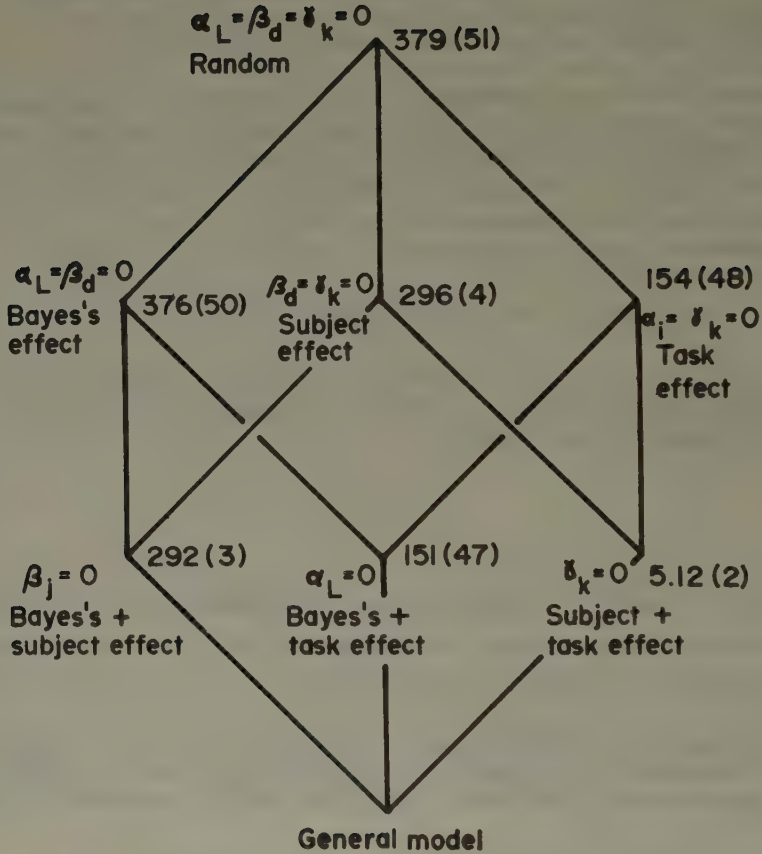


FIG. 5: Lattice of hypotheses associated with Doctor's Trilemma study. At each node the appropriate value of the Chi-squared test statistic is shown with, in brackets, the associated number of degrees of freedom.

the one marked by the asterisk. Thus we must conclude from this study that there are significant subject and task effects but that there is no evidence of a 'Bayes's effect' on subjective performance, though the shortfall of the test quantity 5.12 from the critical value 5.99 is perhaps small enough to encourage the undertaking of more and larger studies.

7.2 Statistician's Syndrome. In this analysis three groups of subjects, 56 professional statisticians, 11 second-year statistics students and 9 clinical consultants, were each presented with five tasks involving the differential diagnosis of three types based on six quantitative features. The complete simulated training set of 36 cases, 12 of each type was given, together with information that for each of these equally prevalent types the distributions of the features were normally distributed, the three mean vectors and covariance matrices also being given to the subjects. These details of the training set and the five new cases are to be found in Aitchison and Kay (1973). For each group there are highly significant subject and task effects. The extensive subject variability can be easily illustrated by Figure 3 which shows in terms of triangular coordinates the inferential statement of the 56 statisticians for one of the cases. If, for a given task, group g has inferential statements which are $L_2(\mu_g, \Sigma_g)$ distributed then interest is in testing such hypotheses as $\Sigma_1 = \Sigma_2 = \Sigma_3$ and $\mu_1 = \mu_2 = \mu_3$. We have made such comparisons between our three groups for each task with the following results. For only one of the tasks is there no significant difference between the groups. For all the remaining four tasks there are significant differences between the mean vectors, highly significant at 0.1 per cent for three of these tasks, though for only one task is there a significant difference, at 5 per cent, between the covariance matrices.

Thus there is evidence, not only that subjects within a group vary but that there can be significant differences between the performance of different groups' subjects. Since a normative statistical system, such as predictive statistical diagnosis can for each task supply a normative inferential statement translatable into logratio value μ we can test whether a group's μ_g is significantly different from μ . For all three groups and all five tasks there are significant differences of group means from the normative value.

ACKNOWLEDGEMENTS

I am grateful to all the subjects, in particular the many students and colleagues in the Universities of Glasgow and Hong

Kong and the school children taking part in Open Days, who, by their patience, willingness and enthusiasm in allowing themselves to be compared with 'normative' inference, have provided me with the experience and stimulus to enthuse on the usefulness of performance analysis.

REFERENCES

- Aitchison, J. (1974). Hippocratic inference. *Bulletin of the Institute of Mathematics and its Applications*, 10, 48-53.
- Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, 62, 547-554.
- Aitchison, J. (1981). Distributions on the simplex for the analysis of neutrality. In *Statistical Distributions in Scientific Work*, C. Taillie, G. P. Patil and B. Baldessari, eds. D. Reidel Publishing Company, Dordrecht, Holland. Vol. 4, pp. 147-156.
- Aitchison, J. and Dunsmore, I. R. (1976). *Statistical Prediction Analysis*. Cambridge University Press.
- Aitchison, J., Habbema, J. D. F. and Kay, J. W. (1977). A critical comparison of two methods of statistical discrimination. *Applied Statistics*, 26, 15-25.
- Aitchison, J. and Kay, J. W. (1973). A diagnostic competition. *Bulletin of the Institute of Mathematics and its Applications*, 9, 382.
- Aitchison, J. and Kay, J. W. (1975). Principles, practice and performance in decision making in clinical medicine. In *The Role and Effectiveness of Theories of Decision in Practice*, D. J. White and K. C. Bowen, eds. Hodder and Stoughton, London.
- Aitchison, J., Moore, M. F., West, S. A. and Taylor, T. R. (1973). Consistency of treatment allocation in thyrotoxicosis. *Quarterly Journal of Medicine*, 42, 573-583.
- Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, 67, 261-272.
- Davidson, D., Suppes, P. and Siegel, S. (1957). *Decision-making: An Experimental Approach*. Stanford University Press.
- Edwards, W. and Phillips, L. D. (1964). Emerging technologies for making decisions. In *Human Judgments and Optimality*, pp. 360-401. Wiley, New York.
- Goldberg, L. R. (1970). Man versus model of man: a rationale, plus some evidence for a method of improving on clinical inferences. *Psychological Bulletin*, 73, 422-432.
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, 14, 107-114.
- Kay, J. W. (1976). Diagnosis and performance. Ph.D. dissertation, University of Glasgow.
- Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory*. Dover, New York.

- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 525-540.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27, 986-1005.
- Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72, 346-354.
- Raiffa, H. (1968). *Decision Analysis*. Addison-Wesley, Reading, Massachusetts.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 and 623-656.
- Taylor, T. R., Aitchison, J. and McGirr, E. M. (1971). Doctors as decision makers: a computer-assisted study of diagnosis as a cognitive skill. *British Medical Journal*, 3, 35-40.

[Received May 1980]

EVERY BODY HAS ITS MOMENTS

ERNEST G. ENNS*, P. FRANK EHLERS and STEIN STUHR

Department of Mathematics & Statistics
The University of Calgary
Calgary, Alberta, Canada T2N 1N4

SUMMARY. The distributions of ray and secant lengths through arbitrary convex bodies for five different randomness measures has previously been obtained. This article concentrates on the ray and secant moment relationships that exist among these various measures. In particular various moments depend only on the volume and surface content of the n -dimensional convex body. Specific moment relationships have been obtained for regular convex bodies which include all n -dimensional regular polyhedra. The result will be illustrated by applications.

KEY WORDS. Geometrical probability, convexity, random secants.

1. INTRODUCTION

Random paths through convex bodies have attracted considerable interest recently. An introduction to the subject can be found in Kendall and Moran (1963) with an updating by Moran (1966, 1969). More recent work can be found in Coleman (1969), Miles (1969) and Kingman (1969). Extensive references to applications may be found in Kellerer (1971).

This article will deal with moment relationships of rays and secants that occur under various types of randomness. The

* The author acknowledges the support of the National Science and Engineering Research Council of Canada.

formalism will be an extension of that developed in Enns and Ehlers (1978, 1980).

A secant is a straight line segment traversing a convex body K . A ray is a line segment originating within the body K and terminating on its surface. The lengths of these rays and secants is now determined by their generation mechanism. Five different methods of generation will be considered, namely:

- ν -randomness: A point within K and a direction, each with independent uniform distributions, can be used to define a ray and a secant.
- μ -randomness: A secant is defined by a direction θ and by its intersection P with an $(n-1)$ -dimensional hyperplane through some fixed origin and normal to θ . The point P and direction θ have independent uniform distributions.
- λ -randomness: Two points are chosen independently within K , each with a uniform distribution. These points define three random lengths, namely a ray, a secant and the distance between the two points.
- α -randomness: Two points are selected independently, one on the surface of K and the other inside K . Both points are obtained from uniform distributions over their respective domains. These points define a ray and a secant.
- γ -randomness: A point on the surface of K and a direction into K , each with independent uniform distributions can define a secant of K .

The distributions of these ray and secant lengths have been obtained in the above references and are summarized in Appendix I.

It will be shown that for a special class of convex bodies, which we will call regular, moments of ray and secant lengths can be related for all five types of randomness. This could be useful in determining the type of randomness which has generated a sample of secants which can be measured. This will be illustrated in the applications.

Every body has its moments! If we've experienced the right moments, how much do we know about the body? This question, too, is answered in our applications.

2. OVERLAP VOLUMES AND SURFACES

The distributions in this paper (see Appendix I) have been obtained from or are written in terms of overlap volumes and surfaces of the convex region under consideration. Let $S(\cdot)$ and $V(\cdot)$ represent the surface content and volume of (\cdot) respectively. Let $K(\ell, \theta)$ represent the convex region K translated a distance ℓ in direction θ . If $E_\theta(\cdot)$ denotes the mean of (\cdot) when uniformly averaged over direction, then define:

$$\Omega(\ell) = E_\theta[V(K \cap K(\ell, \theta))]/V(K)$$

$$\omega(\ell) = E_\theta[S(K \cap K(\ell, \theta))]/S(K).$$

These represent the normalized average overlap volume and surface content respectively of K with its translated self.

If K' is a shell of uniform width h about K , then in Enns and Ehlers (1980), it was found that

$$\begin{aligned}\tilde{\Omega}_K(\ell) &\equiv \lim_{h \rightarrow 0} [\Omega_{K \cup K'}(\ell) - \Omega_K(\ell)]/h \\ &= \frac{S(K)}{V(K)} [\omega_K(\ell) - \Omega_K(\ell)].\end{aligned}\quad (1)$$

The subscript K will be dropped whenever there is no ambiguity about the region under consideration.

From the form of the distributions in Appendix I, it is evident that one can obtain two sets of moment relationships. One set relates ν , μ and λ randomness moments, the second set α and γ randomness moments. Equation (1) will be used to relate these two sets.

3. MOMENT RELATIONSHIPS FOR ARBITRARY K

Let $E_r(X^m)$ denote the expected value of the m th moment of the random variable X under r randomness where $r \in [\nu, \mu, \lambda, \alpha, \gamma]$ and $X \in [R, L, T]$ represents either a ray, secant or segment respectively.

The distributions of R under ν -randomness and L under γ -randomness yield the following moment relationships:

$$E_V(R^m) = m \int_0^\infty \ell^{m-1} \Omega(\ell) d\ell \quad (2)$$

$$E_Y(L^m) = m \int_0^\infty \ell^{m-1} \omega(\ell) d\ell. \quad (3)$$

Moments of all other random variables can be expressed in terms of the above basic moments. The two sets of relationships, where "n" represents the dimension throughout, are given by:

$$\begin{aligned} E_V(R^{m+n}) &= E_V(L^{m+n}) / (m+n+1) = \frac{E_\lambda(L^{m+n+1}) C_{n-1} S(K)}{n(m+n+1) C_n V(K)} \\ &= E_\lambda(R^m) V(K) / C_n = \frac{E_\lambda(L^m) (n+1) V(K)}{C_n (n+1)} \\ &= \frac{E_\lambda(T^m) (n+m) V(K)}{n C_n}. \end{aligned} \quad (4)$$

$$\text{and } E_Y(L^{m+n}) = \frac{2(n+m) V(K)}{n C_n} E_\alpha(R^m) = 2 V(K) E_\alpha(L^m) / C_n. \quad (5)$$

The question now becomes, can these two sets of moment relationships be connected. A special class of convex bodies will be considered in Section 4 where this connection can be explicitly obtained.

3.1 Moments Which Depend Only on the Surface Content and Volume of K. The following results have appeared in Enns and Ehlers (1978, 1980):

$$\int_0^\infty \Omega''(\ell) d\ell = -\Omega'(0) = \frac{C_{n-1} S(K)}{n C_n V(K)},$$

$$\int_0^\infty \ell^{n-1} \Omega(\ell) d\ell = \frac{V(K)}{n C_n}, \quad \int_0^\infty \ell^{n-1} \omega(\ell) d\ell = \frac{2V(K)}{n C_n}.$$

Combining these results with (2), (3), (4) and (5), one obtains the following moment relationships which depend only on the volume and surface content of K:

$$E_V(R^n) = (E_\lambda(R^{-n}))^{-1} = V(K) / C_n \quad (6)$$

$$E_Y(L^n) = (E_\lambda(L^{-n}))^{-1} = (n+1) V(K) / C_n \quad (7)$$

$$E_{\gamma}(L^n) = (E_{\alpha}(L^{-n}))^{-1} = 2 V(K)/C_n \quad (8)$$

$$E_{\mu}(L) = E_{\nu}(L^{-1})^{-1} = \frac{n C_n V(K)}{C_{n-1} S(K)} \quad (9)$$

$$E_{\mu}(L^{n+1}) = (E_{\lambda}(L^{-n-1}))^{-1} = \frac{n(n+1)}{C_{n-1}} \frac{(V(K))^2}{S(K)}. \quad (10)$$

Equations (7), (9) and (10) have also appeared explicitly or implicitly in Hadwiger (1950), Kingman (1969) and Miles (1969).

4. REGULAR CONVEX BODIES

A regular convex body K is defined as a region whose surface content can be obtained by differentiating the equation of the volume with respect to a single variable. Examples include the n -sphere, n -cube, rhombus, regular polygons and regular polyhedra of all dimensions. The definition also includes more complicated regions such as a cylinder with a hemispherical cap on both ends, namely a sausage.

If K is regular, then $V(K)$ can be expressed in terms of a variable b , such that $dV(K)/db = S(K)$. In this case the volume of $K \cup K'$ is $V(K)$ with $b \rightarrow b+h$. Consequently, for regular K , (1) may be written as

$$\tilde{\Omega}(\ell) = d\Omega(\ell)/db = [\omega(\ell) - \Omega(\ell)]S(K)/V(K). \quad (11)$$

Multiplying (2) by ℓ^{m-1} and integrating over ℓ , one obtains the following differential equation relating moments:

$$\frac{d}{db} E_{\nu}(R^m) = [E_{\gamma}(L^m) - E_{\nu}(R^m)] \frac{d}{db} \ln V(K). \quad (12)$$

If we let $E_{\gamma}(L^m) = (A+1) E_{\nu}(R^m)$, then (12) becomes:

$$\frac{d}{db} E_{\nu}(R^m) = A E_{\nu}(R^m) \frac{d}{db} \ln V(K). \quad (13)$$

Clearly A must be a dimensionless quantity. Thus, either A is a function of " n " and " m " only, or it additionally depends in a dimensionless manner on the length b . The latter is possible only if the volume of K is a function of two or more lengths. An example of this occurs when K is a sausage of radius b and length h . The surface area can be obtained by differentiating the volume with respect to b only, hence the

sausage is a regular convex body. In this case A could be a function of the dimensionless quantity (b/h) .

If $V(K)$ is a function of the length b only, then of necessity A cannot be a function of b . In this case the solution of (13) is:

$$E_V(R^m) = C(V(K))^A.$$

In order that the dimensionality of this expression be correct, we must have $A = m/n$. Hence for regular convex bodies defined by a single dimensional quantity:

$$E_\gamma(L^m) = \left(\frac{n+m}{n}\right) E_V(R^m). \quad (14)$$

The corresponding relation between secants under v and γ randomness is:

$$E_v(L^m) = \frac{n(m+1)}{n+m} E_\gamma(L^m).$$

Ehlers (1972) discovered this relationship for the square and n -sphere. Coleman (1973) gives it for the square and cube. It should be noted that (6) and (8) imply that (14) is true for *any* convex body when $m = n$ ($n = \text{dimension}$).

4.1 Further Regularity Conditions. The previous section can be extended by considering regions K where more than one dimensional parameter is necessary. An example is a rectangle which is defined by a length and a width. We will restrict ourselves to regions where the volume $V(K)$ can be expressed in terms of the variables b_1, b_2, \dots, b_r ; namely

$$V(K) = f(b_1, b_2, \dots, b_r)$$

such that

$$\lim_{h \rightarrow 0} \left[\frac{V(K \cup K') - V(K)}{h} \right] = \sum_{i=1}^r \frac{\partial f(b_1, \dots, b_r)}{\partial b_i} = S(K).$$

Regions so defined include the rectangle but not the ellipse.

Relation (1) can now be written as:

$$\sum_{i=1}^r \frac{\partial \Omega(\ell)}{\partial b_i} = \frac{S(K)}{V(K)} [\omega(\ell) - \Omega(\ell)]. \quad (15)$$

For example, the rectangle $2b_1 \times 2b_2$, leads to

$$\omega(\ell) = \Omega(\ell) + \frac{b_1 b_2}{(b_1 + b_2)} \left[\frac{\partial \Omega}{\partial b_1} + \frac{\partial \Omega}{\partial b_2} \right].$$

The γ -random moments of secant length may thus be obtained directly from the ν -random moments as

$$E_{\gamma}(L^m) = E_{\nu}(R^m) + \frac{b_1 b_2}{b_1 + b_2} \left[\frac{\partial}{\partial b_1} E_{\nu}(R^m) + \frac{\partial}{\partial b_2} E_{\nu}(R^m) \right],$$

which is easily seen to be consistent with Coleman's (1973) results. Similar results hold for the n -dimensional box and, more generally, for the n -dimensional rhombus.

5. APPLICATIONS

Kellerer (1971) lists extensive references of the applications of random chord length transversals of convex bodies. The problems are broadly of three types.

1. If the randomness is known and the type of body is known (i.e. triangle, ellipsoid, etc.) but the size of the body is unknown, then this can be estimated via the moments. This is self-evident and will not be elaborated upon.
2. If the size and shape of the convex body is known, but the type of randomness is unknown, then this can be deduced in most cases from the secant length data. This is illustrated in Section 5.1 for the triangle.
3. If nothing is known about the body, then Section 3.1 can be used to deduce the volume and surface area of the body provided one can generate a suitable type of randomness. One method whereby this can be achieved is if K is imbedded in a spherical container K' with a radius much larger than the maximum diameter of K . Coleman (1969) has shown that γ -random secants generated through K' become μ -random secants within K . If the μ -random secants can be measured then using (9) and (10) one can deduce the volume and surface area of K . In particular if $\bar{\ell}$ and $\bar{\ell}^4$ are the observed μ -secant first and fourth moments, then the volume and surface area in 3-dimensions can be estimated by:

$$V(K) = \pi \bar{\ell}^4 / (3\bar{\ell}) \quad \text{and} \quad S(K) = 4\pi \bar{\ell}^4 / (3(\bar{\ell})^2).$$

5.1 *Deducing the Randomness of Secants Through an Equilateral Triangle.* Appendix II states the new results $\Omega(\ell)$, $\omega(\ell)$ and the γ -randomness moments for the equilateral triangle. Via Sections 3 and 4 all other randomness moments may be obtained. The means and variances are:

$$\begin{array}{ll} E_{\gamma}(L) = 0.5245 h & \text{var}_{\gamma}(L) = 0.0924 h^2 \\ E_{\mu}(L) = 0.5236 h & \text{var}_{\mu}(L) = 0.0920 h^2 \\ E_{\nu}(L) = 0.6994 h & \text{var}_{\nu}(L) = 0.0622 h^2 \\ E_{\alpha}(L) = 0.7898 h & \text{var}_{\alpha}(L) = 0.0430 h^2 \\ E_{\lambda}(L) = 0.8425 h & \text{var}_{\lambda}(L) = 0.0310 h^2 \end{array}$$

By examining a sample of secants with unknown randomness, one could apply the standard classification procedures, such as in Anderson (1958) to determine the appropriate randomness. Obviously it will be difficult to distinguish between γ and μ -randomness for the triangle.

REFERENCES

- Coleman, R. (1969). Random paths through convex bodies. *Journal of Applied Probability*, 6, 430-441.
- Coleman, R. (1973). Random paths through rectangles and cubes. *Metallography*, 6, 103-114.
- Ehlers, P. F. (1972). *Particle number fluctuations and geometrical probability*. Ph.D. Thesis, University of British Columbia.
- Enns, E. G. and Ehlers, P. F. (1978). Random paths through a convex region. *Journal of Applied Probability*, 15, 144-152.
- Enns, E. G. and Ehlers, P. F. (1980). Random paths originating within a convex region and terminating on its surface. *Australian Journal of Statistics*, to appear in April.
- Enns, E. G. and Ehlers, P. F. (1980). Random secants of a convex body generated by surface randomness. *Journal of Applied Probability*, to appear.
- Hadwiger, H. (1950). Neue Integralrelationen für Eikörperpaare. *Acta Scientiarum Mathematicarum*, 13, 252-257.
- Kellerer, A. M. (1971). Considerations on the random transversal of convex bodies and solutions for general cylinders. *Radiation Research*, 47, 359-376.
- Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical Probability*. Griffin, London.
- Kingman, J. F. C. (1969). Random secants of a convex body. *Journal of Applied Probability*, 6, 660-672.

- Miles, R. E. (1969). Poisson flats in Euclidean spaces. Part I: Finite numbers of random uniform flats. *Advances in Applied Probability*, 1, 211-237.
- Moran, P. A. P. (1966). A note on recent research in geometrical probability. *Journal of Applied Probability*, 3, 453-463.
- Moran, P. A. P. (1969). A second note on recent research in geometrical probability. *Advances in Applied Probability*, 1, 73-89.

APPENDIX I

RAY AND SECANT LENGTH PROBABILITY DENSITY FUNCTIONS FOR ARBITRARY K.

Let R , L and T be random variables representing the ray length, secant length and the distance between two points respectively. Let the density functions be written as $f_{X;r}(\ell)$ where $X \in [R, L, T]$ and r is the randomness, namely $r \in [\nu, \mu, \lambda, \alpha, \gamma]$.

$$f_{R;\nu}(\ell) = -\Omega'(\ell)$$

$$f_{L;\nu}(\ell) = \ell \Omega''(\ell)$$

$$f_{L;\mu}(\ell) = -\Omega''(\ell)/\Omega'(0)$$

$$= n C_n V(K) \Omega''(\ell)/[C_{n-1} S(K)]$$

$$f_{R;\lambda}(\ell) = -C_n \ell^n \Omega'(\ell)/V(K)$$

$$f_{L;\lambda}(\ell) = C_n \ell^{n+1} \Omega''(\ell)/[(n+1) V(K)]$$

$$f_{T;\lambda}(\ell) = n C_n \ell^{n-1} \Omega(\ell)/V(K)$$

$$f_{R;\alpha}(\ell) = n C_n \ell^{n-1} \omega(\ell)/2V(K)$$

$$f_{L;\alpha}(\ell) = -C_n \ell^n \omega'(\ell)/2V(K)$$

$$f_{L;\gamma}(\ell) = -\omega'(\ell)$$

where primes denote differentiation with respect to ℓ and C_n is the volume of unit n -sphere.

APPENDIX II

The overlap area and circumference of an equilateral triangle of height h are respectively:

$$\frac{\pi}{3} \Omega(\ell) = \begin{cases} \frac{\pi}{3} - 2\rho + \left(\frac{\pi}{6} + \frac{\sqrt{3}}{4} \right) \rho^2 & \text{if } 0 \leq \rho \leq 1 \\ \frac{\pi}{3} - 2\rho + \left(\frac{\pi}{6} + \frac{\sqrt{3}}{4} \right) \rho^2 + 3\sqrt{\rho^2 - 1} - (2 + \rho^2) \sec^{-1} \rho & \text{if } 1 \leq \rho \leq 2/\sqrt{3} \end{cases}$$

$$\frac{\pi}{3} \omega(\ell) = \begin{cases} \frac{\pi}{3} - \rho & \text{if } 0 \leq \rho \leq 1 \\ \frac{\pi}{3} - \rho + 2\sqrt{\rho^2 - 1} - 2 \sec^{-1} \rho & \text{if } 1 \leq \rho \leq 2/\sqrt{3} . \end{cases}$$

where $\rho = \ell/h$. The γ -randomness moments are:

$$E_Y(L^m) = h^m A_m / (m+1)$$

where $A_1 = 3(\ln 3)/\pi$, $A_2 = 2\sqrt{3}/\pi$
 $A_3 = (3 \ln 3 + 4)/2\pi$, $A_4 = 20/(3\sqrt{3} \pi)$

$$A_{k+2} = \frac{kA_k}{k+1} + \frac{3}{\pi(k+1)} \left(\frac{2}{\sqrt{3}} \right)^{k+1}, \quad k = 0, 1, 2, \dots$$

[Received May 1980]

SOME DISTRIBUTIONS IN THE THEORY OF GRAPHS

MICHAEL CAPOBIANCO

St. John's University
Staten Island, New York 10301 USA

SUMMARY. In certain types of populations, one is interested in the relationships among the individuals, e.g., animals and plants in a food web, persons in a social group, cities linked by various transportation routes, etc. These can be modeled by graphs or digraphs (directed graphs). Trying to infer something about these structures from a subgraph (sample) constitutes statistical inference in graphs (what we have called "stagra-phics"). There are many distributions which occur in graph theory which could be useful in inference. This paper is a survey of some of these distributions, specifically, degree distributions, path length distributions, distance distributions, cycle distributions, triad distributions, and distributions connected with pairs of stars. Various properties of these distributions will be discussed, as will their connections with inference. Several open problems will be presented.

KEY WORDS. graphs, digraphs, statistics, distributions.

1. INTRODUCTION

1.1 Graphs. A graph G , is non-empty set, V , of elements we call the *points* or *vertices* of G , together with a set, E , of unordered distinct pairs of distinct points of G . The elements of E are called *lines* or *edges* of G . If $v, w \in V$, and $(v, w) \in E$, then we say that v and w are *adjacent*. If we denote the lines (v, w) by e , say, then we say that v and w are *incident* with e . The cardinality of the set of vertices,

$|V|$, denoted by N , is called the *order* of the graph, and $|E|$, denoted by R , is called the *size* of the graph.

Graphs can be represented by diagrams in which the points are dots, and adjacencies are indicated by drawing line segments between two dots representing adjacent points. Figure 1 shows a representation of the graph $V = \{1, 2, 3, 4, 5, 6, 7\}$, $E = \{(2, 3), (4, 6), (4, 7), (5, 6), (6, 7)\}$.

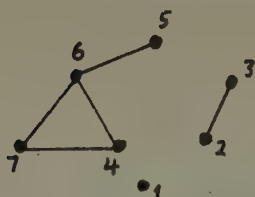


Fig. 1

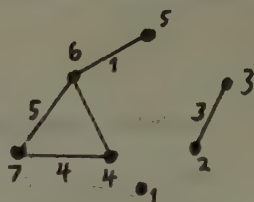


Fig. 2

They can also be represented by certain $(0,1)$ -matrices. We define an *adjacency matrix* of a graph G by labeling the points with integers from 1 to N as, for example, in Figure 1. The $N \times N$ adjacency matrix $A = [a_{ij}]$ has all its entries equal to zero except that $a_{ij} = 1$ if points labeled i and j are adjacent in G . Thus for the graph represented in Figure 1, the adjacency matrix is

$$\begin{array}{c}
 \begin{array}{ccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
 \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}
 \end{array}
 \end{array}
 .$$

If we label the lines as well as the points, say as in Figure 2, we can define the $R \times N$ *incidence matrix* $B = [b_{ij}]$ whose entries are $b_{ij} = 1$ if point labeled j is incident with line labeled i in G . Thus, B for the graph of Figure 2 is

$$\begin{array}{c}
 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \\
 \begin{bmatrix}
 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 1
 \end{bmatrix}
 \end{array}$$

1.2 Digraphs. A *digraph* (directed graph), D , is a non-empty set, V , of elements called points or vertices of D , together with a set, E , of *ordered* distinct pairs of distinct points of D . The elements of E are called lines or *arcs* of D . If $v, w \in V$, and $(v, w) \in E$, then we say that v is *adjacent to* w (or w is *adjacent from* v). $|V|$ and $|E|$ are called the order and size of D and are denoted N and R , respectively. Digraphs can be represented by diagrams in which points are dots, and arcs are arrows. For example, the digraph $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{(1, 2), (2, 3), (3, 4), (4, 1), (1, 4), (8, 7), (3, 5)\}$ is shown in Figure 3.

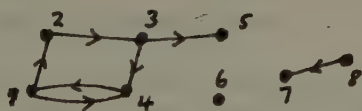


Fig. 3

The adjacency matrix $A = [a_{ij}]$ for a digraph is $N \times N$ and has all entries 0 except that $a_{ij} = 1$ if point i is adjacent to point j in D . Thus, for the digraph of Figure 3,

$$A = \begin{array}{c}
 \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\
 \begin{bmatrix}
 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
 \end{bmatrix}
 \end{array}$$

1.3 The Purpose of This Paper. In certain problems, we are interested in the relations among individuals in a population, e.g., the food web of an animal and plant community, a social group where the relation could be acquaintance, the hierarchical structure of authority in an organization, a communication or transportation network, etc. Such populations, referred to by us as "populations with structure" (Capobianco, 1970) can be effectively modeled by graphs and digraphs. The points represent the individuals, and the lines or arcs display the relations.

Our main research interest is in statistical inference in graphs, or "stagraphics", which is the word we coined for it, i.e., inferring something about the graph (structured population) by observing a sample (a subgraph) from it. See Frank (1980b,c) for surveys of this subject. O. Frank is one of the originators of work in this field. It was in connection with problems of this type that we came across several properties of graphs which can be thought of as distributions, and it is to these that this working paper is devoted. It is intended as a survey, which, it is hoped, will stimulate further research in this area. Hence, few proofs will be given, but many concepts are explained and illustrated, several theorems are stated, and some unsolved problems are presented.

2. DEGREE DISTRIBUTIONS

2.1 Degrees. The *degree* of a point, v , in a graph, denoted $d(v)$, is the number of lines incident with v . For example, in the graph of Figure 1, $d(1) = 0$, $d(2) = d(3) = d(5) = 1$, $d(4) = d(7) = 2$, and $d(6) = 3$. In a digraph, any point v has an *indegree* and an *outdegree*. These are denoted $id(v)$ and $od(v)$, respectively, and are defined as the number of vertices adjacent *to* v , and the number of vertices adjacent *from* v . The (total) degree, $td(v)$, of a point v is given by

$$td(v) = id(v) + od(v). \quad (1)$$

For the digraph of Figure 3, we have $id(1) = id(2) = id(3) = id(5) = id(7) = 1$, $id(4) = 2$, $id(6) = id(8) = 0$, $od(1) = od(3) = 2$, $od(2) = od(4) = od(8) = 1$, $od(5) = od(7) = od(6) = 0$.

We present next an easy, but important, result known as the "first theorem of graph theory", although its corollary sometimes goes by that name. We include the proofs since they are not difficult, and will help introduce the reader to the "flavor" of the subject.

Theorem 2.1. For any graph G ,

$$\sum_v d(v) = 2R \quad (2)$$

where the sum is taken over all the vertices of G , and R , as usual, denotes the size of G .

Proof. One of the quickest proofs is obtained by using the incidence matrix. We count the number of 1's in this matrix in two different ways, first by rows, then by columns. The former yields the value $2R$ since there are R rows, and each one has two 1's. The latter yields $\sum_v d(v)$ since the sum of each

column is obviously the degree of the point corresponding to the column.

Corollary. In any graph, the number of points of odd degree is even.

These results also hold for the total degrees in a digraph. In addition, we have the following:

Theorem 2.2: In any digraph.

$$\sum_v id(v) = \sum_v od(v) \quad (3)$$

where the sums are taken over all points in the digraph.

Proof. Using the adjacency matrix provides an easy proof. We count the number of 1's in two ways. Row-wise, we obtain the right side of (3), and column-wise we obtain the left side.

An important problem in graph theory is to determine when a sequence of integers is "graphical", i.e., when there exists a graph having points whose degrees are the integers of the given sequence. It is not sufficient that the integers sum to an even number. For example, there do not exist graphs having points with degrees 1, 2, 3, or degrees 1, 1, 1, 3, 4. Hakimi (1962) gives necessary and sufficient conditions for a graphical sequence, while Berge (1962, p. 86) gives a corresponding result for digraphs.

2.2 Degree Distributions of Graphs. The degree distribution of a graph is a set of numbers p_i , $i = 0, 1, 2, \dots, N-1$ where N is the order of the graph, and p_i is the proportion of vertices having degree i . Clearly, the p_i 's form a probability distribution.

It will be convenient to illustrate degree distributions by means of "spaced-out" histograms such as in Figure 4, which illustrates the degree distribution for the graph of Figure 1.

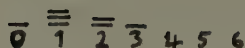


Fig. 4



Fig. 5

It is important to note that two graphs can be very different and have the same degree distribution. An example is shown in Figure 5.

Frank (1978), using induced subgraph sampling, gave unbiased estimates for the degree distribution, paying special attention to p_0 . In induced subgraph sampling, n points are chosen at random and one observes all of the adjacencies among them.

The *complement* of a graph G , denoted \bar{G} , is a graph whose points are the points of G , but in which two points are adjacent if they are not adjacent in G , and not adjacent if they are adjacent in G . In other words, to complement a graph we remove all lines and insert lines where there were none. The following results are immediate:

Theorem 2.3. If p_i , $i = 0, 1, \dots, N-1$, and \bar{p}_i , $i = 0, 1, \dots, N-1$, are the degree distributions of G and \bar{G} respectively, then $\bar{p}_i = p_{N-1-i}$.

In other words, the degree distributions of complements are reflections of each other about a central vertical axis. Figure 6 shows the complement of the graph of Figure 1 and its degree distribution.

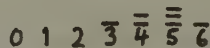
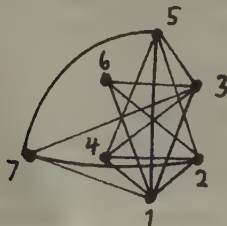


Fig. 6

We may ask if there are graphs for which $\bar{p}_i = p_i$, $i = 0, 1, \dots, N-1$, i.e., graphs which have *symmetric* degree distributions. It is well known that there are graphs which are *self-complementary*, i.e., they are the same as their complements. Three of these are shown in Figure 7. Obviously, they have symmetric degree distributions. Do there exist graphs with symmetric degree distributions which are not self-complementary? Yes! An example is shown in Figure 8.

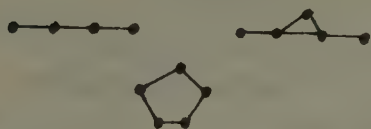


Fig. 7

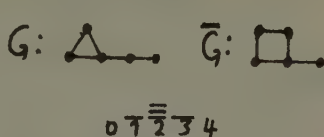


Fig. 8

Theorem 2.4. No graph has a nondegenerate uniform degree distribution.

In fact, any graph having more than one point must have at least two vertices having the same degree. It is, however, possible to have a uniform distribution over part of the range of the degrees, such as, for example, in Figure 5.

We define the *mean degree* μ_d of a graph as the mean of its degree distribution, i.e.,

$$\mu_d = \sum_{i=0}^{N-1} i p_i.$$

Since this is $\sum_v d(v)/N$, we have from Theorem 2.1, that

$\mu_d = 2R/N$. Thus, all graphs of the same order and size have the same mean degree. Of course, there are many graphs of different order and size with the same mean degree, e.g., the two shown in Figure 9. From Theorem 2.3, the mean degree of the complement of G , $\bar{\mu}_d$, is $N-1-\mu_d$.

The variance of the degree distribution is defined by

$$\sigma_d^2 = \sum_{i=0}^{N-1} i^2 p_i - \mu_d^2.$$

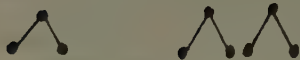


Fig. 9

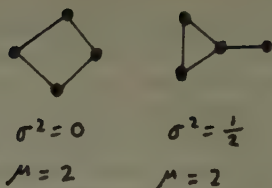


Fig. 10

It is easy to get examples of graphs with equal means and different variances. See Figure 10. It is also easy to prove that $\sigma_d^2 = \bar{\sigma}_d^2$, where $\bar{\sigma}_d^2$ is the variance of the degree distribution of the complement. The variance is important in the comparison of several estimators of graph size (Capobianco and Frank, 1979).

Cook (1979) has proved several theorems dealing with variances of degrees and sums of squares of degrees. In order to understand them, we require a few definitions from graph theory. A *path* in a graph is a sequence of distinct points v_1, v_2, \dots, v_k in which v_i and v_{i+1} are adjacent for $i = 1, 2, \dots, k-1$. If v_1 and v_k are the same, then the path is called a *cycle*. A graph without cycles is said to be *acyclic*.

Theorem 2.5 (Cook). For any acyclic graph.

$$\sum_v [d(v)]^2 \leq N^2 - N$$

with equality if and only if the graph is the *star* shown in Figure 11.

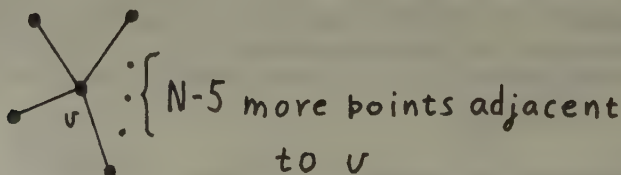


Fig. 11

A graph is *planar* if it can be embedded in the plane. Loosely speaking, this means that it can be drawn in the plane without any lines intersecting except at the vertices.

Theorem 2.6 (Cook). For a planar graph,

$$\sigma_d \leq \sqrt{2N} + O\left(\frac{1}{\sqrt{2N}}\right).$$

Since the complement of any planar graph of order at least 9 is nonplanar (Battle *et al.*, 1962) this inequality holds for infinitely many non-planar graphs as well.

The minimum degree in a graph is denoted δ , and the maximum, Δ . Hence the range of the degree distribution is $\Delta - \delta$. The minimum degree is important in problems dealing with the connectivity of a graph. A graph is *connected* if there is a path between any two points. The minimum number of points which must be removed (along with their incident lines of course) in order to disconnect the graph, or reduce it to a single point, is called the *connectivity* of the graph and denoted κ . A well-known theorem of graph theory states that $\kappa \leq \delta$. Capobianco (1972) used this to estimate κ .

Much work has been done on degree distributions of random graphs. See Karóński (1978) and Robinson and Schwenk (1975).

2.3 Degree Distributions of Digraphs. In the case of digraphs, the degree distribution is bivariate because we have indegrees and outdegrees. We will picture these distributions as shown in Figure 12, which is the degree distribution of the digraph of Figure 3. For these distributions we use μ_{id} , μ_{od} , σ_{id}^2 , σ_{od}^2 and ρ to denote the means, variances, and correlation coefficient of in and out degrees, respectively. It follows from (3) that for any digraph $\mu_{id} = \mu_{od} = R/N$.

The investigation of other properties of these distributions is a wide open area. How does one interpret ρ for instance? What is the relationship between the two variances?

3. DISTANCE DISTRIBUTIONS

3.1 Introduction. The *length* of a path in a graph is the number of edges in the path. The *distance* between two points, u , v , of a graph denoted by $d(u,v)$, is the length of a shortest path between u and v . The *diameter* of a graph is the maximum distance between two of its points.

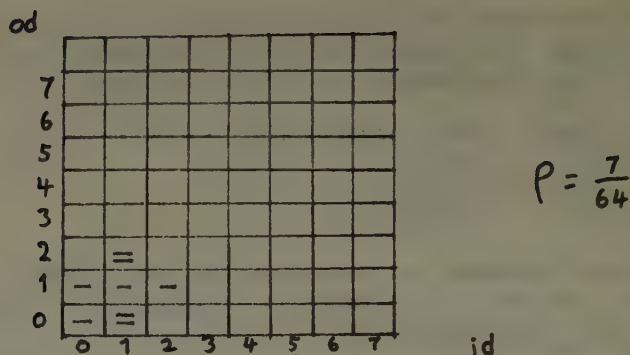


Fig. 12

The *distance distribution* of a graph G is a set of numbers q_i , $i = 0, 1, 2, \dots, N-1$ where q_0 is the proportion of pairs of points G which have no path between them (distance infinity), and q_i , $i = 1, 2, \dots, N-1$, is the proportion of pairs of points (u, v) of G such that $d(u, v) = i$, i.e.,

$$q_0 = \frac{\text{number of pairs of points with no path between them}}{\binom{N}{2}}$$

$$q_i = \frac{\text{number of points } (u, v) \text{ such that } d(u, v) = i}{\binom{N}{2}}$$

for $i = 1, 2, \dots, N-1$.

Again, we will display distance distributions by means of spaced-out histograms. Figure 13 gives the distribution for the graph of Figure 1.

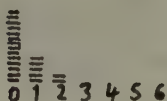


Fig. 13

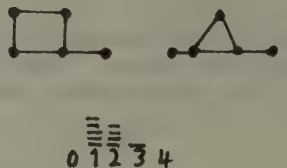


Fig. 14

The following result is immediate.

Theorem 3.1: (i) $q_0 = 0$ iff G is connected.

$$(ii) \quad q_1 = 2R/(N(N-1)) = \sum_v d(v)/(N(N-1)).$$

(iii) If $q_0 \neq 0$, the diameter of G is infinite. If $q_0 = 0$, the diameter of G is the largest i for which $q_i \neq 0$.

The following result is easy to prove by induction on N . It was found useful in testing the hypothesis that a graph is connected (Capobianco, 1980b).

Theorem 3.2. If G is connected, then

$$q_1 + q_2 \geq \frac{4N-6}{N(N-1)}.$$

It should be noted that two graphs can be different but have the same distance distribution. Figure 14 shows an example.

For connected graphs, we define the mean distance μ_D and variance of the distance distribution σ_D^2 as expected, namely,

$$\mu_D = \sum_{i=0}^{N-1} i q_i, \quad \sigma_D^2 = \sum_{i=0}^{N-1} i^2 q_i - \mu_D^2.$$

This mean distance is a measure of how "compact" a graph is. A graph is called *complete* if every pair of vertices is adjacent. Complete graphs are the only ones with $\mu_D = 1$. The opposite extreme is a graph which is a path. In this case, $\mu_D = (N+1)/3$. It is also straightforward to compute $\sigma_D^2 = (N+1)(N-2)/18$, for a path. Is this the upper bound on variance?

Three points of a graph are said to be *collinear* if they can be labeled u, v, w in such a way that $d(u,v) + d(v,w) = d(u,w)$. We define the *collinearity* of a graph as the proportion of collinear triples of points in the graph, i.e.,

$$CL = \frac{\text{numbers of collinear triples}}{\binom{N}{3}}$$

where CL denotes collinearity (Capobianco, 1980a). The following result relates collinearity with mean distance.

Theorem 3.3. For any connected graph G of order at least three,

$$CL \geq 3(\mu_D - 1)/(N-2).$$

Thus, collinearity is also related to compactness. Equality holds above for graphs which are *geodetic*. This means that they have exactly one shortest path (geodesic) between any two points. We define the *geodeticity*, g , of a non-complete graph by

$$g = 3(\mu_D - 1)/(N-2)CL$$

and take $g = 1$ for a complete graph. This is a measure of how geodetic a graph is which can be estimated by observing only distances!

3.2 Uniform Distance Distributions. There are graphs with distance distributions which are uniform over part of the range. Figure 15 shows two examples.

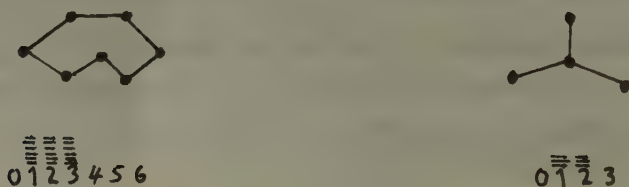


Fig. 15

In fact, we have the following:

Theorem 3.4. Any cycle of odd order has distance distribution

$$q_i = \begin{cases} \frac{2}{N-1} & i = 1, 2, \dots, \frac{N-1}{2} \\ 0 & i = \frac{N+1}{2}, \frac{N+3}{2}, \dots, N-1 \end{cases}$$

Theorem 3.5. Any cycle of even order has distance distribution

$$q_i = \begin{cases} \frac{2}{N-1} & i = 1, 2, \dots, \frac{N}{2} - 1 \\ \frac{1}{N-1} & i = \frac{N}{2} \\ 0 & i = \frac{N+2}{2}, \frac{N+4}{2}, \dots, N-1 \end{cases}$$

A graph is called *bipartite* if the set of vertices can be partitioned into two sets of nonadjacent points. If all adjacencies possible between the two sets of points are present, then the graph is called a *complete bipartite* graph. We can show the following:

Theorem 3.6. Complete bipartite graphs in which the two sets of vertices have cardinalities

$$\binom{K+1}{2} \text{ and } \binom{K+2}{2} \quad k = 1, 2, \dots$$

have distance distribution $q_1 = q_2 = \frac{1}{2}$.

The graph in Figure 16 has distance distribution $q_1 = q_2 = q_3 = 1/3$, so that Theorems 3.4 and 3.6 do not yield all graphs with "uniform" distance distributions. In fact, the graph in Figure 17 is not complete bipartite but has distance distribution $q_1 = q_2 = \frac{1}{2}$!

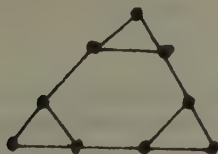


Fig. 16



Fig. 17

Because the existence of a path of length k implies the existence of one of lengths $k-1$, $k-2$, and so on to one of length 1, one may get the impression that a distance distribution must be "non-increasing." This is false. Figure 18 gives examples for which the mode is central, and my colleague, F. Buckley, found the graph of Figure 19 for which the mode is the diameter!!

Investigating the properties of these distributions appears to be a fruitful field of research. What relations can we derive between μ_D and σ_D^2 ? Between σ_D^2 and the diameter? How can we characterize graphs with uniform distance distributions?

3.3 Miscellaneous Questions. Work has been done in the area of distance distributions of random graphs. See Karónski (1978), Meir and Moon (1978,1975), Moon (1971).

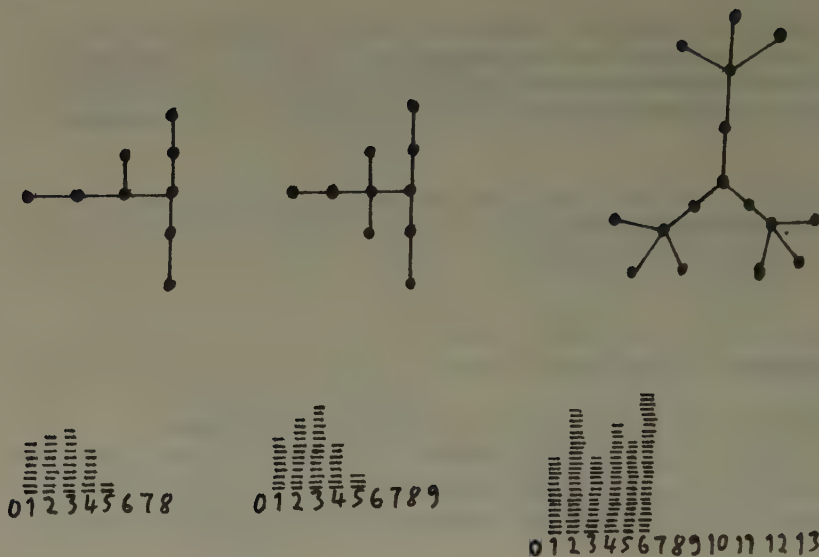


Fig. 18

Fig. 19

For digraphs, distances must be directed. If two points do not have a dipath (directed path) between them, they are at infinite distance. Figure 20 shows an example. No work at all, to our knowledge, has been done on distance distributions of digraphs. We suspect this area is rich with strange results.

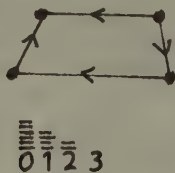


Fig. 20

4. CYCLE DISTRIBUTIONS

4.1 Introduction. The length of a cycle is the number of edges, or points, in it. A cycle of length k is called a k -cycle. Defining the cycle distribution of a graph presents rather serious problems. It is easy enough to tally up the frequencies of 3-cycles, 4-cycles, etc., but deciding what relative frequency, i.e., probability, should mean causes severe difficulties. One

might say it is simply the number of k -cycles divided by the total number of cycles. However, doesn't it seem more reasonable that the probability of a k -cycle, r_k , say, should be

$$\frac{\text{the number of } k\text{-cycles}}{\binom{N}{k}} ?$$

These r_k 's do not, of course, sum to unity! Perhaps we should use

$$r_k = \frac{\text{the number of } k\text{-cycles}}{2^N - N - \binom{N}{2} - 1}.$$

Here, the denominator is the total number of sets of points of cardinality, at least three. But this is surely wrong. We should use as numerator the number of these sets which contain a k -cycle. But then what do we do when one of them contains more than one k -cycle, as, for example, in the graph of Figure 21?

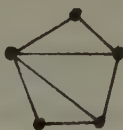


Fig. 21

5. OTHER DISTRIBUTIONS

Holland and Leinhardt, who are among the originators of statistical inference in graphs, have worked with the distribution of the triads in digraphs (1970, 1975). These are the 16 possible subdigraphs that can be formed using three points. See also Frank (1978a,b) and Wasserman (1977)

Capobianco (1978, 1979) used distributions based on the interaction of pairs of points in a digraph. Thinking of a digraph as a model for a sociometric choice pattern, two cases were considered, fixed choice and variable choice. In the former, each person (point) was restricted to naming exactly three friends. In the latter, all friends in the group were to be named. Hence, in the fixed choice scheme, each point has outdegree three, while in the variable case a point can have any outdegree from zero to $N-1$.

In the fixed choice case, examination of how pairs of points interact yields only ten different underlying configurations. These are shown in Figure 22. The encircled points in the figure are the points of the pair in question. The distribution of the configurations under randomness was computed so that the distribution of a sample of pairs could be compared to that expected under randomness. This technique seems to be effective for obtaining indications of certain characteristics of the population, clustering, symmetry, reciprocal choices, etc.

In the variable choice situation, a trivariate distribution was derived for a pair of points. The random variables were j , the number of lines between the points of the pair, K , the number of points chosen by both points of the pair, and M , the total number of points which were chosen by any one point of the pair.

This distribution is given by the probability function

$$f(j) = \frac{(N-2)!}{K! (N-2-K-m)! m!} 2^m p_1^{m+2K} p_0^{2(N-2-K)-m} \quad (1)$$

where $f(0) = p_{00}$, $f(1) = 2p$, $f(2) = p_{11}$; $p_1 = p + p_{11}$, $p_0 = p + p_{00}$. The correlation matrix is

$$\begin{bmatrix} 1 & 0 & 0 \\ & 1 & \frac{2p_1^2}{\sqrt{2p_1(2-p_0)(p_0^2+p_1^2)}} \\ & & 1 \end{bmatrix}$$

Plotting K vs. m using three different symbols for the values of j is a technique which seems to provide information about the population digraph.

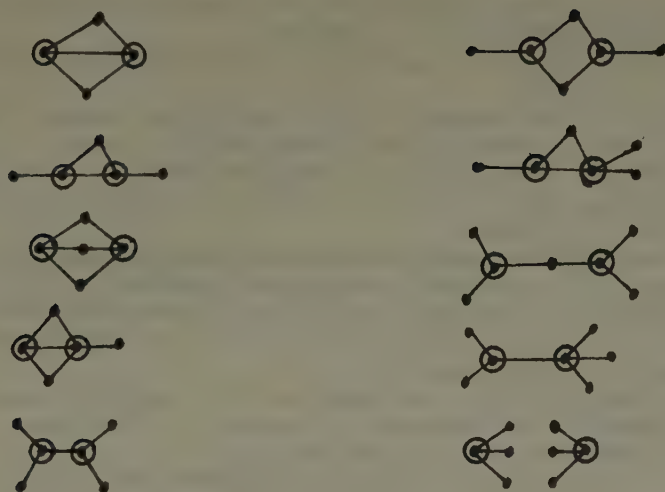


Fig. 22

REFERENCES

- Battle, J., Harary, F., and Kodama, Y. (1962). Every planar graph with nine points has a non-planar complement. *Bulletin of the American Mathematical Society*, 68, 569-571.
- Berge, C. (1962). *The Theory of Graphs and Its Applications*. Wiley, New York.
- Capobianco, M. (1970). Statistical inference in finite populations having structure. *Transactions of the New York Academy of Sciences*, 32, 401-413.
- Capobianco, M. (1972). Estimating the connectivity of a graph. In *Graph Theory and Applications*, Y. Alavi, D. Lick, A. White, eds. Springer-Verlag, New York.
- Capobianco, M. (1978). Some problems in staghaphics. Presented at the Joint Statistical Meetings, San Diego.
- Capobianco, M. (1979). Exploring graphs using pairs of stars in a sample. Presented at the Twelfth Annual Meeting of European Statisticians, Varna.
- Capobianco, M. (1980a). Collinearity, distance, and connectedness of graphs. To appear.
- Capobianco, M. (1980b). Introduction to statistical inference in graphs (staghaphics) and its applications. Under revision for *SIAM Review*.
- Capobianco, M. and Frank, O. (1979). Comparison of statistical graph size estimators. Submitted.
- Capobianco, M. and Molluzzo, J. (1978). *Examples and Counter-examples in Graph Theory*. Elsevier, North Holland, New York.

- Cook, R. (1979). Vertex degrees of planar graphs. *Journal of Combinatorial Theory B*, 26, 337-345.
- Frank, O. (1978). Sampling and estimation in large social networks. 1, 91-101.
- Frank, O. (1979). Estimating a graph from triad counts. *Journal of Statistical Computing and Simulation*, 9, 31-46.
- Frank, O. (1980). Estimating the number of vertices of different degrees in a graph. *Journal of Statistical Planning and Inference*. To appear.
- Frank, O. (1980b). Sampling and inference in a population graph. *International Statistical Review*, 48, 33-41.
- Frank, O. (1980c). A survey of statistical methods for graph and analysis. *Sociological Methodology 1981*, S. Leinhardt, ed. Jossey-Bass, San Francisco.
- Hakimi, S. (1962). On the realizability of a set of integers as the degrees of the vertices of a graph. *SIAM Journal of Applied Mathematics*, 10, 496-506.
- Holland, P. and Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 70, 492-513.
- Holland, P. and Leinhardt, S. (1975). Local structure in social networks. In *Sociological Methodology 1976*, D. R. Heise, ed. Jossey-Bass, San Francisco.
- Karóński, M. (1978). *A Review of Random Graphs*. Uniwersytet Im. Adama Mickiewicza W Poznaniu, Instytut Matematyki Poznań.
- Meir, A. and Moon, J. (1975). Climbing certain types of rooted trees, I. In *Proceedings of the Fifth British Combinatorial Conference*, Aberdeen, 461-469.
- Meir, A. and Moon, J. (1978a). Climbing certain types of rooted trees, II. *Acta Mathematica Hungarica Academy of Sciences*, 31, 43-54.
- Meir, A. and Moon, J. (1978b). On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30, 997-1015.
- Moon, J. (1971). The lengths of limbs of random trees. *Mathematika*, 18, 78-81.
- Robinson, R. and Schwenk, A. (1975). The distribution of degrees in a large random tree. *Discrete Mathematics*, 12, 359-372.
- Wasserman, S. (1977). Random directed graph distribution and the triad census in social networks. *Journal of Mathematical Sociology*, 5, 61-86.
- Wright, E. (1974). Large cycles in large labelled graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 78, 7-17.
- Wright, E. (1977). Large cycles in large labeled graphs II. *Mathematical Proceedings of the Cambridge Philosophical Society*, 81, 1-2.

[Received May 1980. Revised October 1980]

COGRADUATION BETWEEN STATISTICAL DISTRIBUTIONS AND ITS APPLICATIONS — A GENERAL REVIEW

ADOLFO GILI

Instituto di Statistica
Universita di Bologna
Bologna, Italy

1. THE NOTION OF COGRADUATION AND ITS IMPLICATIONS IN THE THOUGHT OF CORRADO GINI

The notion of cograduation is very elementary and it seems almost impossible that on it a whole body of doctrine may be built, which, owing to its organic unity, to the extent of its subject and its implications was and still is one of the fundamental contributions to that trend of thought called "Italian statistical school".

If we consider a double statistical distribution we shall call *correspondent* the quantities actually associated in the same classification unit, whereas we shall define *cograduated* those occupying the same position in the relative classifications. With regard to the notion of cograduation several statistical indexes have been put forward to measure how much, in the mean, the connotation of correspondence between quantities differed from that of cograduation.

Every proposed index corresponded obviously to specific purposes of investigation determined each time by the needs of the concrete research.¹ It is only in 1914, however, that Corrado Gini clearly focused the possibility for the above mentioned notions to be considered as the basis of a complete and unitary theoretic tissue which will be called "concordance theory". The work written in 1914 draws only a few premises based on the notion and the measure of the dissimilarity.² In the following works, which cover a period of time roughly corresponding to the First World War, Gini was developing all the implications following from the cited work, in order to define the concordance

theory³. The reading of these works is not easy and it becomes particularly difficult when there is the problem of seizing the leading thread which link them up and when it is necessary to explain the essential passages of this way.

On this observation it is possible to base the necessity, that Salvemini felt in the forties⁴, of resuming this subject, in order to connect or better to clear up the connection among the cells of a mosaic. It is the author of this work who, only two years ago, deemed it opportune to study this subject once again, to give a contribution for the explanation of the conceptual premises from which the whole theory by Gini results.⁵ If all that is opinable from a theoretical point of view, it is also true that both Salvemini and the author of this work have considered only one of the concrete meanings that the whole theory by Gini may take.

Now we are going to examine in short the foundations of this theory. We shall start by pointing out the distinction among characters, attributes that cannot be ordered and classifications. The first ones refer to quantitative aspects of the real and may be divided in enumerable and measurable. The second ones refer to qualitative connotations and each of them differs in the same way from all the others; therefore it is not possible to fix a logical succession. The classifications, which refer only to the places concerning the entities at issue, can relate only to characters or to attributes that can not be ordered. For these it is also possible to develop a theory equivalent to that outlined for the characters.

If the notion of cograduation is considered in a strict sense it regards only the comparisons between classifications. But it is just through the concept of dissimilarity by Gini that this notion can play an essential role also as to the comparison between characters and, owing to a few propositions that we are not going to consider now, as to the comparison among attributes that cannot be ordered.⁶

As a matter of fact if we dwell for a while upon the aspects which are strictly connected with the notion of cograduated entities, or if you like, of cograduation, meaning by that the proceeding to cograduate two successions having the same number of terms, it is possible to come spontaneously and rapidly to the determination of the cograduation indexes.⁷ But these imply other concepts which involve the whole theory by Gini. Consequently it is necessary to mention them as well. These concepts will permit then to conceive a cograduation index as an index of concordance between classifications. On the other hand the concept of concordance represents another main point of this paper.

To introduce this notion it is necessary to dwell upon the characters owing to their chief importance and because what will be said about the theory referring to the attributes that cannot be ordered and the classifications, might always result as a particular case of what we have already argued about the characters.

Here the distinction between correspondent entities and cograduated entities comes back as a main point of the paper. Because of our simple treatment of this subject we shall always consider them as pertaining enumerable characters and therefore we shall speak again of quantities.

As a matter of fact by the word "*discordance*" Gini means also the difference between correspondent quantities and we can mark them with 1M the mean arithmetic value of discordances, that is to say of the differences between correspondent quantities considered in absolute value and, likewise, with 2M the mean arithmetic value of the squares of the discordance. When instead of the discordance we consider the differences in absolute value between cograduated quantities or the squares of the differences, it is possible to come, by calculating the mean, respectively to the simple dissimilarity index and to the square of the quadratic index.⁸

It is now interesting to consider the meaning that the value 0 of the dissimilarity index (either simple or quadratic) may assume, not only to explain the meaning implied in this particular value of the index, but also because another fundamental concept for the concordance theory arises from it.

Then we start observing that to the value 0 of the indexes at issue corresponds the notion of similar simple statistical distributions (i.e. defined by equal quantities to which equal relative frequencies correspond in an orderly way). If this concept is supposed sound for each pair of simple subordinate statistical distributions forming a double statistical distribution and also for the two simple statistical distributions forming the double statistical one, it is possible to come to the definition of the association frequencies in the hypothesis of independence. We denote with 1M_0 the mean arithmetic value of differences in absolute value between quantities in the hypothesis of independence, and with 2M_0 the mean arithmetic value of the squares of these quantities in the same hypothesis.

It is now necessary to fix a criterion to verify if between the simple statistical distributions forming the double statistical distribution there exist or not any concordance, reserving to ourselves the right of explaining later on the sense of the expression "non-concordance". In that respect

the problem is very simple, if we refer to discordance squares, whereas it turns out rather complex if we consider the absolute values.⁹ Therefore we are simply going to develop quadratic discordances. We shall say together with Gini that there exist a concordance when the quantity 2M results inferior to the quantity 2M_0 . If these quantities have equal value we shall say again, agreeing with Gini, that between the simple statistical distributions forming the double statistical distribution there is *indifference*, as this concept has got a meaning which is less restrictive than that of independence. The equation at issue in fact does not necessarily imply that these association frequencies appearing in 2M are the same ones characterizing the definition of 2M_0 . In this paragraph we do not consider the hypothesis where the quantity 2M turns out superior to the quantity 2M_0 . As a matter of fact in this case we usually speak of *discordance*, but this is, in our opinion, a concept that cannot be specified without introducing the notion of contrary distribution. Here the paper is too premature, as we are going to deal with this matter in Section 2. We remember that the difference $^2M_0 - ^2M$ characterizes the nature of the criterion of α -concordance by Gini. It is always sufficient to refer to it when squares of differences are used.

To define a concordance index (it does not matter if considered in a strict sense or in a broad sense)¹⁰ it is necessary to relate the above mentioned difference to the maximum value it can assume, since only in this case the index will vary in a definite interval. Here another problem arises, as regards the way of interpreting the maximum of concordance.

For simplicity's sake we shall face it referring to the notion of concordance in a strict sense. Given the simple statistical distributions forming the double one we are sure to have a maximum of concordance (in a strict sense), when the mean value 2M coincides with the square of the quadratic index of dissimilarity between those distributions. It is clear, however, that the concordance may be perfect only when these latter distributions turn out similar (or if you like equal, since in this context the notion is not restrictive). In the concrete meanings two cases may occur: in the former case the simple statistical distributions forming the double one are logically antecedent to the association of the quantities at issue in the classification units; in the latter they are consequent.¹¹ It is possible to understand then that in the former case the maximum of concordance must be the one we have already pointed out and therefore we shall speak of relative maximum of concordance. In the latter case the simple statistical distributions are not, conceptually, conditions regarding the possible ways according to which the associations may take place. Therefore they can be changed so as to coincide, i.e. so that the

quadratic index of dissimilarity may become null; we shall speak then of absolute maximum of concordance (in a strict sense).

With reference to the relative maximum of concordance¹² it is possible to outline the definition of the quadratic indexes of homophily, whereas as regards the absolute one it is possible to outline the definition of the correlation indexes.

As far as the attributes that cannot be ordered are concerned we shall define indexes of *attraction* the concordance indexes based on the notion of relative maximum and *resemblance* indexes those referring to the notion of absolute maximum. Even if it is not necessary we like to observe that we refer to quadratic indexes in this context.¹³

Gini's theory does not consider the distinction between relative maximum and absolute maximum with regard to quadratic indexes of cograduation, that is to say, as we have already mentioned, with regard to indexes pertaining to the concordance between classifications. As a matter of fact if to each entity given at the beginning, we assign a different place in the classification, it can be easily deducted that the relative maximum and the absolute maximum of concordance coincide and therefore Gini speaks only of the quadratic index of cograduation.¹⁴

Recently, however, one of our collaborators felt the necessity, justified by a substantial research, of considering sets of quantities with an equal classification place and consequently of introducing the frequency of each set. This has led him, also in terms of quadratic cograduation indexes, to take back Gini's theory as to the distinction between relative maximum and absolute maximum of concordance, by the introduction of a quadratic index of cograduation that makes use of the relative maximum of concordance.¹⁵ With this contribution the picture that Gini had drawn in the far 1916 referring to the Concordance Indexes, is completed.

2. - QUADRATIC INDEXES OF CONCORDANCE

In the preceding pages we have observed that the quantitative foundation of the concordance measure lies in the difference $2M_0 - 2M$. We have also observed that it is possible to attain a concordance index by relating the above mentioned quantity to the maximum value that it can assume, which, in the hypothesis of relative maximum turns out $2M_0 - 2D^2$, where $2D^2$ is the square of the quadratic dissimilarity index, that notoriously measures the minimum possible distance between two double statistical distributions forming the double statistical

distribution. The quotient between the first and the second difference we have mentioned characterizes the quadratic homophily index. It is obvious that this construction has got a sense if we speak of concordance in the strict meaning, that is supposing such result $2M_0 - 2M > 0$. The problem that consequently arises, following the α -criterion of concordance¹⁶, which consists in analysing and interpreting the sign of the difference $2M_0 - 2M$, regards the arguments which are to be developed when this sign is negative. We can state without any preplexity that in this case between the simple statistical distributions forming the double statistical distribution there is discordance. Now it must be carefully analysed on what conceptual foundation the maximum value, that this negative difference may assume, must be determined. The purpose of this is to construct a quadratic index of discordance.

This problem is actually examined and solved by Gini by using the notion of cograduation.¹⁷ If we indicate with $2M_2$ the mean arithmetic value of the squares of differences between contragraduated quantities, this quantity is known to result a maximum and therefore the structure of the quadratic index of discordance cannot be formally equal to that of the quadratic index of concordance (in a strict sense). Gini then follows a proceeding whose spontaneity we are not able to understand. But the conceptual question is another one; it consists in pointing out that by operating in this way it is not possible to use the notion of dissimilarity index, which is certainly one of the main points of the whole theory. On the other hand the notion of contragraduation is in itself a formal variant of that of cograduation and hence arises clearly the need of introducing a concept that may give a theoretical foundation, consistent with the given premises, to the discordance indexes on which the quadratic index has still got to operate. On the other hand it was Gini himself¹⁸ who felt clearly the logical necessity of introducing a notion that we have only used to overcome the less brilliant aspects of the construction we have already referred to¹⁹. It is the notion of *contrary*. If we indicate with α_i the generic modality of the simple statistical distribution A , component of the double statistical distribution (A, B) , we shall call with α the *simple statistical distribution contrary* of the distribution A , whose modalities result defined by the following relations: $\alpha_i + \alpha_{m-i+1} = K_A$, being m the number of the modalities of the distribution A and K_A constant of addition; once this constant has been defined the distribution α is marked. In fact as regards the frequencies it is almost unnecessary to remark that $n_i^{(A)} = n_{m-i+1}^{(\alpha)}$. In the concrete the problem may be solved by assigning the mean arithmetic value of the distribution α , as it is clear that $M_{1\alpha} - M_{1A} = K_A$. As a matter of fact here one

should distinguish between complementary characters and non-complementary characters, since for the first ones there is no problem of choice as to the definition of the contrary distribution. In any case when from the modalities of the character we pass on to the deviates (from the arithmetic means) it results: $\alpha_i - \alpha_{m-i+1} = 0$ and therefore for the deviates there is not the problem of any subjective choice in order to define the contrary distribution.

We shall call, agreeing with Gini, inverse quadratic index of dissimilarity²⁰ the dissimilarity index built between a simple statistical distribution and the contrary of the other simple statistical distribution forming the double statistical distribution.

If we make use of the notion of inverse quadratic index of dissimilarity the quadratic index of heterophily results to be logically constructed on the same way as the quadratic index of homophily.

From the conceptual point of view, however, the question is more subtle and it regards the autonomy of the concept of discordance. Indeed we have observed that the notion of contragraduation is not characterized in itself, in our opinion, but only as a formal variant of that of cograduation. It seems to us logically meaningful to underline that this critical interpretation may be also suitable to the notion of discordance, in the sense that it appears as a formal variant of the notion of concordance and in this case is involved the concordance between a simple statistical distribution and the simple statistical distribution contrary to the other, forming the double statistical distribution.

It is almost superfluous to remember that the quadratic index of heterophily results univocally determined by choosing indifferently the contrary distribution of one of the two simple statistical distributions forming the double one.²¹ In the substantial scientific research the distinction that Gini, however, clears punctually as to the comparison between intensities, deviates and standardized variates²², may find a concrete legitimation; even if it is true that Fortunati maintains clearly how the use of the comparison between deviates or between standardized variates must actually imply the logical justification of the comparison between intensities.²³ On the other hand the three meanings of the mentioned quantities result connected, in geometric terms, by elementary linear relations.²⁴ Welcoming this distinction commonly accepted in the applications, we want to remember that the quadratic indexes of heterophily and homophily keep equal formal structure either in the case of the concordance measure between deviates or between standardized

variates. Gini comments this result as a negative aspect of the quadratic indexes at issue. In our opinion, however, the problem is to distinguish, on one hand, between the formal structure of an index and the possibility of interpreting the index itself in different ways, on the other to think again carefully over the observation made by Gini, in the light of the argument developed by Fortunati.

We are going to deal now with the quadratic indexes of concordance which refer to the notion of absolute maximum of concordance.

If we consider the concordance in the broad sense, the above mentioned indexes, that we call, agreeing with Gini, of correlation, must be able to oscillate in the closed interval $[-1, +1]$ assuming the value -1 in the hypothesis of absolute maximum discordance, and the value $+1$ in that of absolute maximum of concordance.

Set in the above mentioned terms the problem concerning the construction of a quadratic correlation index, it is clear that the quadratic index of dissimilarity between the simple statistical distribution A and B (or that between the respective deviates or between the respective standardized variates) is to result null and the quadratic dissimilarity indexes built with regard to the distributions α and B and to the distributions A and β are to result simultaneously null too. In order that this might happen it is necessary for the four above mentioned distributions to turn out symmetrical and coincident.²⁵ Gini expresses himself like that. It is not a matter of verifying if these distributions are symmetrical and coincident or not. It is logically possible to define a double statistical distribution whose simple statistical distributions are symmetrical and coincident, and coincident with the respective contrary distribution as well. This aim can be easily achieved by defining a simple statistical distribution as arithmetic mean of the distributions A , B , α and β .²⁶ For the construction of the denominator of the quadratic indexes of correlation we make use of a distribution which is symmetrical and coincident with the respective contrary distribution; from this fact it follows that both the direct quadratic indexes and the inverse quadratic indexes of dissimilarity result simultaneously null. The use of the plural in the last notes may seem wrong, as it is obvious that if one pair of indexes is obtained with reference to the relative maximum of concordance, with regard to the absolute one only one index results to be built. This plural, however, concerns another aspect of the problem that we are going to clear up immediately. It is in fact easy to demonstrate that three quadratic indexes of correlation may be obtained; they refer respectively to the measure of concordance between intensities, deviates and standardized variates.²⁷

Among the pieces of this great mosaic we find again the coefficient of correlation by Bravais Peason in the quadratic index of correlation between standardized variates.

We would like to stop our exposition with the last sentence, but this would mean to believe that there are conclusive moments in scientific research.

As a matter of fact if we consider the structure of the quadratic index of correlation between intensities we realize²⁸ that this is not equal to +1 in the hypothesis of absolute maximum of concordance, that is to say when the distribution A coincides with the distribution B and likewise the distribution α coincides with the distribution β .²⁹

If we suppose $M_{1\alpha} = M_{1A}$ and also $M_{1\beta} = M_{1B}$ this value can not be found. But if these conditions are imposed, the structure of the index given by Gini which contains the four mean arithmetic values we have just remembered, is no longer justified. It is not clear how this index may be changed, or better, how the measure of the concordance among intensities may be formulated, so that it is possible to fix an index that avoids the above mentioned troubles.

On the other hand when Fortunati³⁰ suggests to define the constant of addition as twice the arithmetic mean of the simple statistical distribution assigned (and that is to avoid any arbitrary choice in the definition of the contrary distribution), he points out, implicitly, a way of solving the question in hand. It is obviously taken for granted that the formal structure of the index will appear with a less general meaning than that given by Gini.

It is also true that Gini himself distinguishes between characters which admit a complementary character and characters for which the complementary one is not determined or determinable.³¹

In respect to a character admitting the complementary one Gini argues about the spontaneity of assuming this latter as the contrary of that given at the beginning. In this case, however, we are not able to accept the methodical indication given by Fortunati.

NOTES

1. C. Spearman, "*Footrule*" for measuring correlation, in "The British Journal of Psychology", II Vol., 1906. This index has been found again by K. Pearson, in a context which shows a different kind of substantial problems in the work On

further methods for determining correlation, in "Draper's Co Research Memoires Biometric Series" II, 1907. A different cograduation index is that suggested by Kendall in the work: *A New Measure of Rank Correlation*, in "Biometrika", XXX Vol. 1938. These indexes, but not the simple cograduation index by Gini which we are going to mention in the note 7), result to be particular cases of a cograduation index introduced by M. Badaloni with A. Rizzi in the work *Contributi alla "cluster analysis"*, in "Metron", XXX Vol., no. I-4, 1972.

2. C. Gini, *Di una misura della dissoiglianza tra due gruppi di quantita e delle sue applicazioni allo studio delle relazioni statistiche*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti", Academic year 1914-15, LXXIV Tome, II part, Venice, 1915.
3. C. Gini, *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti", Academic year 1914-15, LXXIV Tome, II Part, Venice 1915; Id.Id., *Nuovi contributi alla teoria delle relazioni statistiche*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti," Academic year 1914-15, LXXIV Tome, II Part, Venice 1915; Id.Id., *Sul criterio di concordanza tra due caratteri*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti", Academic year 1915-16, LXXV Tome, II Part, Venice 1916.; Id. Id. *Indici di concordanza*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti", Academic year 1915-16, LXXV Tome, II Part, Venice 1916; Id. Id., *Delle relazioni tra le intensita cograduate di due caratteri*, in "Atti del R. Istituto Veneto di Scienze, Lettere ed Arti," Academic year 1916-17, LXXVI Tome, II Part, Venice 1917.
4. T. Salvemini, *Lezioni di Statistica metodologica*, III Part *Le relazioni statistiche*, University of Rome, Academic year 1945-46.
5. A. Gili, *Alcune esplicitazioni sui fondamenti della teoria della concordanza*, in "Statistica", XXXVII Year, no. 3, 1977. Id. Id., *Indici quadratici di concordanza*, Quaderno no. 2 of the "Collana di quaderni" of the Istituto di Statistica - Universita di Bologna, Cooperative Libreria Universitaria Editoriale, Bologna, 1978.
6. C. Gini, *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, cit.; Id. Id., *Nuovi contributi alla teoria delle relazioni statistiche*, cit., Id. Id., *Indici di concordanza*, cit.

7. C. Gini, *Di una misura delle relazioni tra le graduatorie di due caratteri*, appendix of the monography: *Le elezioni generali del 1913 nel Comune di Roma*, Cecchini, Roma, 1914. I like to thank Professor Tommaso Salvemini as he has got a copy of this appendix for me. C. Gini, *Indici di concordanza*, cit. The index at issue is a simple cograduation index and consequently we are not going to deal expressly with in this work; we only remember that it is constructed making use of the criterion of concordance δ (without losing anything from a general point of view).
8. It is well known that on the theory of dissimilarity Gini builds also the connection theory, giving thus the word *dependence* a clear and original meaning. The simple index of connection appears for the first time in the work: *Di una misura della dissomiglianza tra due gruppi di quantita e delle sue applicazioni allo studio delle relazioni statistiche*, cit., the quadratic one in the work *Nuovi contributi alla teoria delle relazioni statistiche*. From the conceptual point of view to those that led Gini to the determination of the homophily indexes. In 1957 Castellano will call these indexes "total indexes of connection" and he will introduce the "global" indexes founding them on a principle equivalent to the one which characterizes the correlation theory according to Gini. See V. Castellano, *Contributi alla teoria della correlazione e della connessione tra due variabili*, in "Metron", XVIII Vol. no. 3-4, 1957. See also: A. Gili, *Aspetti fondamentali della teoria della connessione e concordanza nell'impostazione di Corrado Gini*, publication of the Istituto de Scuola di Statistica dell' Universita di Bologna, Cooperative libreria Universitaria Editoriale, Bologna, 1970.
9. C. Gini, *Sul criterio di concordanza tra due caratteri*, cit.
10. The phrase "concordance in a strict sense" refers to the hypothesis $^2M_0 - ^2M > 0$. The phrase "concordance in a broad sense" or simple "concordance" means concordance in a strict sense or discordance; it is therefore necessary to specify this latter word and to this end we put off the explanation to the 2.2 paragraph.
11. The suggested distinction is somewhat schematic. For more details see C. Gini *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, cit.; A. Gili, *Indici quadratici di concordanza*, cit.

12. In a strict sense.
13. In the memoir *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, cit., Gini comes to the construction of a quadratic resemblance index, that turns out to be the obvious extent, as to the attributes that cannot be ordered, of the quadratic correlation index between variations (referred to characters). In the memoir there is also the notion of *qualities*, *deviations* and *mutations* which refer to the attributes that cannot be ordered like those respectively of *intensities*, *deviations* and *standardized variates* concerning the characters. Later on in the memoir *Nuovi contributi alla teoria delle relazioni statistiche*, cit., Gini introduces the quadratic index of attraction that has got as its logic equivalent the quadratic index of homophily. From Gini's text we can deduce that this index keeps equal structure in respect either with the measure of concordance between qualities, or between deviations or mutations, just as it happens for the quadratic indexes of homophily, as it is mentioned in the 2.3 paragraph. In analogy with the distinction which marks the quadratic indexes of correlation which we are going to deal with in the 2.4 paragraph Gini defined in the memoir *Indici di concordanza*, cit., the quadratic index of resemblance among qualities, the one among deviations and that among mutations.
14. C. Gini, *Indici di concordanza*, cit., the quadratic index of cograduation by Gini coincides with the quadratic index of cograduation introduced by Spearman in the work cited in the note (1). The demonstration is due to Gini himself and appears in the work cited in that note.
15. G. Bettuzzi, *Sulle relazioni fra forme di insediamento a movimento naturale della popolazione emiliana nel periodo 1951-71*, in course of print in the volume "Studi in onore di Paolo Fortunati", Cooperative Libreria Universitaria Editoriale, Bologna. Even if in the present work we do not deal with simple indexes of concordance, we must remember that in 1939 Salvemini introduced a simple index of cograduation with regard to the notion of relative maximum of concordance. See T. Salvemini, *L'indice di cograduazione del Gini nel caso di serie statistiche con ripetizione*, in "Metron", XIII Vol. No. 4, 1939.
16. C. Gini, *Sul criterio di concordanza fra due caratteri*, cit.

17. C. Gini, *Nuovi contributi alla teoria delle relazioni statistiche*, cit.
18. C. Gini, *Nuovi contributi alla teoria delle radazioni statistiche*, cit.
19. A. Gili, *Alcune esplicitazioni sui fondamenti della teoria della concordanza*, cit. Id. Id., *Indici quadratici di concordanza*, cit.
20. C. Gini, *Nuovi contributi alla teoria delle relazioni statistiche* cit.
21. A. Gili *Indici quadratici di concordanza*, cit.
22. C. Gini, *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, cit.
23. P. Fortunati, *Alcune considerazioni sulla impostazione giniana delle misure di concordanza*, in "Atti della XXI Riunione Scientifica della Societa Italiana di Statistica", Bologna, 29th - 30th May 1967.
24. A. Gili, *Sulla proprieta metriche delle costanti statistiche*, in "Statistica", Year XXXII, no. 3, 1972.
25. C. Gini, *Indici di omofilia e di rassomiglianza a loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, cit.; Id. Id., *Indici di concordanza*, cit.
26. T. Salvemini, *Lezioni di Statistica Metodologica*, part III *Le relazioni statistiche*, cit., P. Fortunati, *Alcune considerazioni sulla impostazione giniana delle misure di concordanza*, cit., A. Gili, *Alcune esplicitazioni sui fondamenti della teoria della concordanza*, cit.
27. C. Gini, *Indici di concordanza*, cit.
28. A. Gili, *Una postilla sull'indice quadratico di correlazione tra intensita*, in "Statistica", Year XXXVIII, no. 1, 1978.
29. C. Gini, *Indici di concordanza*, cit. Parallel considerations are worth for what concerns the absolute maximum of discordance and the conditions which let the index assume the value - 1.

30. P. Fortunati, *Alcune considerazioni sulla impostazione giniana delle misure di concordanza*, cit.
31. C. Gini, *Nuovi contributi alla teoria delle relazioni statistiche*, cit.

[Received May 1980]

AUTHOR INDEX

- Afifi, A. A., 7, 11, 23
 Ahmed, A. N., 350, 359
 Ahmad, M., 112, 123
 Ahsanullah, M., 314, 317
 Aitchison, J., 134, 139, 363,
 366, 367, 369, 372, 373,
 374, 377, 378, 380, 383,
 384
 Aiuppa, T. A., 254, 256
 Amato, P., 133
 Amos, D. E., 214
 Anderson, T. W., 147, 154,
 225, 229, 346, 394
 Arjas, E., 284
 Arnold, B. C., 275, 277, 278,
 279, 280, 284
 Ayer, M., 27, 32
 Azzalini, A., 55

 Bahadur, R. R., 167, 220, 229
 Bain, L. J., 171, 173, 179,
 272, 274, 284, 285
 Balkema, A. A., 327, 333
 Bargmann, R. E., 157, 158,
 159, 168, 241, 242,
 245, 247, 252, 256
 Barlow, R. E., 27, 28, 32,
 272, 274, 274, 278,
 280, 282, 285, 310, 315,
 317, 319, 322, 325, 350,
 360
 Barndorff-Nielsen, O., 88, 94,
 108
 Barnett, V., 203, 204, 205,
 207, 209, 213
 Bartlett, M. S., 36, 48, 158,
 220, 229, 262, 270
 Basu, A. P., 169, 170, 172,
 173, 178, 179, 180, 272,
 274, 275, 279, 285, 291,
 306, 335, 336, 337, 338,
 339, 340, 341, 342, 343,
 345, 346
 Bates, G. E., 58, 71
 Battle, J., 413
 Behoodian, J., 112, 123
 Belyayev, Y. K., 13, 22

 Bemis, B. M., 171, 173, 179,
 274, 285
 Benettin, G., 358, 360
 Berge, C., 401, 413
 Berk, B. H., 161, 162, 167
 Berkson, J., 30, 32
 Berlin, B., 346
 Berman, S. M., 337, 346
 Berndt, G. D., 27, 32
 Bhattacharya, C. G., 112, 123,
 Bhattacharya, G. L., 171, 172,
 173, 179
 Bhattacharya, G. K., 182, 191,
 274, 285
 Bildikar, S., 57, 59, 60, 62,
 68, 69, 72, 290, 293, 307
 Billingsley, P., 323, 325
 Birnbaum, Z. W., 336, 346
 Bishop, D. J., 220, 229
 Blischke, W. R., 112, 123
 Block, H., 272, 275, 279, 285
 Block, H. W., 179, 271, 273,
 274, 279, 280, 281, 282,
 283, 284, 285, 291, 307,
 341, 342, 346, 350, 360
 Blom, G., 3, 22
 Blum, J. R., 141, 149, 152,
 153, 154, 169, 180, 352,
 360
 Blumenthal, S., 75, 76, 78,
 80, 81, 82, 83, 84, 85,
 330, 333
 Boswell, M. T., 60, 61, 71,
 111, 123, 139
 Bouver, H., 242, 247, 252, 256
 Bowden, D. C., 112, 124, 125
 Bowman, K. O., 12, 22, 137,
 139, 231, 232, 233, 240
 Box, G. E. P., 158, 220, 224,
 225, 226, 229, 260, 270
 Brain, C. W., 1, 6, 7, 22
 Bray, T. A., 120, 124
 Bresenham, J. E., 27, 33
 Brindley, E. C., 291, 307
 Brodsky, J., 346
 Brown, J. A. C., 134, 139
 Brunk, H. D., 27, 32

- Bryant, J. L., 279, 286
 Bryson, C., 320, 330, 333
 Buchanan, W. B., 282, 283, 286
 Buckley, F., 409

 Capobianco, M., 397, 400, 404, 405, 407, 411, 413
 Sassandro, M., 358, 360
 Chanda, L. C., 35, 36, 43, 47
 Chang, W. C., 112, 123, 125
 Chatfield, C., 57, 59, 71
 Chen, H. J., 8, 23
 Cher, T. C., 93, 94
 Chernoff, H., 44, 48
 182, 191
 Chiang, C. L., 346
 Chieppa, M., 133, 134, 135, 139
 Chung, K. L., 330, 333
 Cinlar, E., 330, 333
 Clark, V. A., 26, 27, 31, 33, 274, 286
 Clifford, P., 346
 Cobb, L., 87
 Coberly, W. A., 112, 125
 Cohen, A., 112, 113, 123, 161, 162, 167, 246, 256
 Coleman, R., 387, 392, 393, 394
 Consul, P. C., 224, 225, 229
 Cook, R., 404, 405, 414
 Cook, W. L., 127
 Corbet, A. S., 57, 72
 Cotterill, D. S., 141, 147, 148, 153, 154, 155
 Cox, D. R., 204, 214, 260, 270, 330, 333
 Cozzolino, J. M., 28, 33
 Crain, B. R., 87, 89, 94
 Cramer, H., 41, 49
 Cramer-Von Mises, 141, 155
 Csörgö, M., 141, 143, 146, 147, 148, 149, 150, 152, 153, 154, 155
 Csörgö, S., 55

 D'Agostino, R. B., 7, 11, 22
 Dahyia, R. C., 28, 33, 80, 82, 83, 84
 Darling, D. A., 147, 154

 David, H. A., 60, 72, 207, 214, 280, 286, 336, 346
 Davidson, D., 366, 384
 Day, N. E., 112, 123, 290, 307
 DeHann, L., 327, 333
 Dempster, A. P., 103, 108, 119, 124
 Desu, M. M., 343, 346
 DeWet, T., 153, 155
 DiCastro, C., 358, 360
 Dick, N. P., 112, 124
 Dishan, M., 328, 334
 Downton, F., 275, 276, 277, 278, 279, 286
 Dugue, D., 147, 155
 Dunsmore, I. R., 377, 384, 385
 Durbin, J., 8, 10, 22, 55, 147, 155
 Dutta, K., 43, 49
 Dystra, R. L., 320, 323, 325

 Ebrahimi, N., 283, 286
 Edwards, C. B., 59, 71
 Edwards, W., 366, 384, 385
 Ehlers, P. F., 387, 388, 389, 390, 392, 394
 Ehrenberg, A. S. C., 57, 59, 71
 Elaudt-Johnson, R., 343, 346
 Elderton, W. P., 242, 247, 256
 El-Neweih, E., 283, 284, 286
 Engelman, L., 224, 229
 Esary, J. D., 277, 279, 281, 282, 283, 286, 314, 317, 320, 325, 350, 360
 Ewing, G. M., 27, 32

 Farlie, D. J. G., 170, 180, 290, 307
 Feller, W., 332, 333
 Fercho, W. W., 13, 22
 Fertig, K. W., 14, 23
 Filliben, J. R., 5, 6, 8, 22
 Fisher, R. A., 57, 72
 Folks, J. L., 160, 167
 Frane, J. W., 224, 229
 Francia, R. S., 4, 23, 56,

- Frank, O., 400, 402, 404, 411, 413, 414
Freschet, M., 290, 307
Freund, J., 273, 274, 275, 279, 280, 286, 290, 293, 307, 342, 346
Friday, D. S., 176, 180, 275, 286, 349
Fryer, J. G., 112, 124, 125
Fu, J. C., 307

Gail, M., 343, 346
Galambos, J., 274, 286, 309, 312, 313, 314, 315, 316, 317, 347
Gardner, L. A., 182, 191
Gaver, D. P., 323, 325
George, E. O., 160, 167
Ghosh, J. K., 336, 337, 338, 339, 340, 341, 343, 345, 346
Ghosh, M., 283, 286
Ghiselli, E. E., 270
Ghurye, S. G., 346
Giesbrecht, F., 134, 139
Gili, A., 415
Gilliland, D. C., 340, 347
Gindler, E. M., 111, 124
Giri, N., 112, 123
Gnandesikan, R., 204, 209, 214
Gnedenko, B. C., 13, 22
Goel, A. L., 77, 84
Goldberg, L. R., 366, 384
Good, I. J., 366, 384
Goodhart, G. J., 57, 59, 71
Goodman, I., 289, 290, 291, 293, 295, 307
Gorodetski, V. V., 36, 48
Gottschalk, V. H., 112, 124
Götze, F., 146, 155
Grannis, G. F., 111, 124
Graser, D. A. S., 307
Graybill, F. A., 38, 49
Gridgeman, N. T., 112, 124
Griffith, W., 283, 286
Gross, A. J., 25, 26, 27, 28, 31, 33, 79, 80, 82, 83, 84, 274, 275, 286
Guldborg, A., 59, 72
Gumbel, E. J., 170, 180, 290, 293, 307, 341, 342, 347
Gupta, R. C., 327, 329, 330, 333
Gurland, J., 59, 71

Habbema, J. D. F., 384
Hadwiger, H., 391, 394
Hahn, G. J., 3, 22
Hajek, J., 170, 180
Hakimi, S., 401, 414
Halfon, E., 245, 256
Hamming, R. W., 244, 256
Hannon, J., 340, 347
Harary, F., 413
Harding, J. F., 112, 124
Harris, C. M., 13, 22
Harris, R., 291, 307
Harter, H. L., 135, 137, 139
Hartley, H. O., 80, 84, 230
Hasselbland, V., 112, 124
Hastings, N. A. J., 132
Hawkes, A. G., 276, 277, 278, 279, 287
Hawkins, D. M., 207, 214
Healy, M. J. R., 204, 209, 214
Hernandez, F., 266, 267, 270
Higgins, J. J., 171, 173, 179, 274, 285, 290, 291, 307
Hill, B.M., 139
Hoeffding, W., 141, 149, 153, 155
Holla, M. S., 79, 84
Holland, P., 411, 414
Hollander, W., 327, 333
Holmes, P. T., 330, 333
Hopf, E., 352, 360
Hosmer, D. W., Jr., 112, 120, 124
Hsieh, H. K., 163, 167, 220, 229
Hsu, C. L., 279, 287
Hurst, H. E., 354

Illinois Department of Transportation, 191

Jagers, P., 330, 333
James, I. R., 112, 124
Jelinski, Z., 76, 83, 84

- Jennrich, R. I., 224, 229
 Hoglekar, A. M., 77, 84
 John, S., 112, 124
 Johnson, N. L., 57, 72, 78, 79,
 84, 85, 132, 134, 139,
 195, 202, 214, 216, 231,
 232, 240, 242, 247, 252,
 254, 256, 257, 272, 273,
 274, 279, 287, 290, 291,
 303, 307
 Johnson, R. A., 171, 172, 173,
 179, 182, 191, 259, 266,
 267, 270, 274, 285
 Jona-Lasinio, G., 358, 360
 Joshi, S. W., 59, 72

 Kac, M., 352, 360
 Kamins, M., 26, 33
 Karónski, M., 405, 409, 414
 Katiyar, R. S., 158, 167
 Katti, S. K., 65, 72
 Kay, J. W., 366, 373, 381,
 383, 384
 Kellerer, A. M., 387, 393,
 394
 Kemp, A. W., 57, 59, 62, 63,
 64, 65, 68, 69, 72
 Kemp, C. D., 68, 69, 72
 Kempthorne, O., 134, 139
 Kendall, D. G., 61, 62,
 72
 Kendall, M. G., 17, 22,
 387, 394
 Kettenring, J. R., 204, 209,
 214
 Katri, C. G., 57, 72, 191,
 226, 230
 Khinchin, A. I., 368, 384
 Kibble, 275
 Kiefer, J., 97, 108, 141,
 149, 152, 153, 154,
 169, 180
 Kiefer, N. M., 119, 120, 124
 Kimeldorf, G., 350, 360
 Kingman, J. F. C., 387, 391,
 394
 Kodama, Y., 413
 Kohberger, 279, 285

 Kolmogorov, A. N., 350, 353,
 356, 360
 Komlós, J., 143, 155
 Kottegoda, N. T., 354, 360
 Kotz, S., 57, 62, 79, 85, 132,
 134, 139, 195, 202, 214,
 216, 217, 247, 252, 254,
 257, 272, 273, 274, 279,
 286, 287, 289, 290, 291,
 293, 295, 303, 307, 313,
 314, 317
 Kounias, E. G., 316, 317
 Krishnaiah, P. R., 224, 229,
 275, 279, 287
 Krishnamoorthy, A. S., 279,
 287
 Krivyakova, E. N., 147, 155
 Kullback, S., 262, 270, 368,
 385
 Kwerel, S. M., 315, 317

 LaBrecque, J., 5, 8, 22
 Laird, N. M., 103, 108, 119,
 124
 Lam, C. F., 275, 286
 Landenna, G., 193
 Langberg, N. A., 343, 347,
 350, 359
 Laud, P. W., 319, 320, 323,
 325
 Laurent, A. G., 333
 Lawrance, A. J., 354, 360
 Lee, A., 112, 125
 Lee, L., 280, 287
 Lee, T. H., 27, 33
 Lehmann, E. L., 44, 48, 88,
 89, 94, 350, 360
 Leibler, R. A., 368, 385
 Leinhardt, S., 411, 414
 Leon, R. V., 350, 359
 Lether, E. M., 243, 257
 Levy, P., 356, 360
 Lewis, T., 203, 204, 205,
 209, 213
 Li, C. C., 77, 85
 Lilliefors, H. W., 9, 23
 Lin, C. C., 13, 22, 158, 167
 Lindley, D. V., 369, 385

- Lindsay, B. G., 95, 96, 98,
105, 109
Lipow, M., 26, 27, 33
Liptak, T., 160, 167
Littell, R. C., 160, 167
Liu, V., 274, 286
Lloyd, D. K., 26, 27, 33
Locke, C., 12, 23
Lockhart, R., 313, 317
Lott, J. A., 111, 124

Macdonald, P. D. M., 112, 124
MacLaren, M. D., 120, 124
Mahfoud, M., 331, 333
Major, P., 143, 144
Maldelbrot, B. B., 354, 355,
356, 359, 361
Malik, H. J., 132
Malkovich, J. F., 7, 11, 23
Mann, H. B., 222, 229
Mann, N. R., 14, 23, 272,
287
Marasini, D., 193
Marcus, R., 76, 78, 80,
82, 83, 84, 85
Mardia, K. V., 242, 243,
247, 257
Marsaglia, G., 120, 124
Marshall, A. W., 169, 170,
172, 180, 274, 275, 277,
278, 279, 281, 282, 283,
284, 286, 290, 291, 293,
307, 308, 314, 317, 320,
325, 341, 342, 347
Martynov, G. V., 147, 155
Mathai, A. M., 158, 167
Mauchly, J. W., 225, 229
Mazumdar, S., 77, 85
Mehrotra, K. G., 172, 180,
275, 287
Meilijson, I., 327, 334
Meir, A., 409, 414
McCloskey, J. W., 61, 72
McGirr, E. M., 366, 367, 369,
372, 385
Michalek, J. E., 172, 180,
275, 287
Miles, R. E., 387, 391, 395
Moeschberger, M. L., 280, 287,
336, 346

Molluzzo, J., 413
Moon, J., 409, 414
Moore, A. A., 134, 139
Moore, M. F., 384
Moore, R. H., 204, 214
Moran, P. A. P., 387, 394,
395
Moranda, P. B., 76, 83, 84,
85
Morgenstern, D., 290, 308
Morrison, D. G., 328, 331,
333, 334
Mucci, R., 316, 317
Mudholkar, G. S., 13, 22,
157, 158, 160, 167,
194, 202, 219
Mumme, D. C., 127
Muth, E. J., 327, 328, 329,
334

Nadas, A., 340, 347
Nagao, H., 157, 158, 167

Nagarsanker, B. N., 225, 226,
229
Narula, S. C., 343, 346
Nath, G. B., 79, 85
Natvig, B., 315, 317
Navin, F. P. D., 313, 317
Nayer, P. P. N., 220, 229
Nelson, W. C., 60, 72
Neyman, J., 72, 220, 230
Nixon, 214

Olkin, I., 169, 170, 172, 180,
274, 275, 278, 279, 284,
287, 290, 293, 308, 341,
342, 347
Oosterhoff, J., 160, 168
Ord, J. K., 111, 123, 242, 257

Page, E. S., 182, 191
Parrish, R. S., 241, 252, 254,
257
Parthasarathy, M., 279, 287
Patankar, V. N., 49

- Patil, G. P., 57, 59, 60, 61,
62, 68, 69, 71, 72, 111,
123, 176, 180, 275, 286,
290, 293, 307
- Paulson, A. S., 207, 214,
277, 278, 279, 285, 287,
288
- Peacock, J. B., 132
- Pearson, E. S., 11, 22, 214,
220, 230, 254, 257
- Pearson, K., 112, 125
- Peliti, L., 358, 360
- Pesarin, F., 51, 55, 56
- Peters, B. C., 112, 125
- Peterson, A. V., Jr., 347
- Pettitt, A. N., 10, 11, 23,
56
- Philippou, N., 62, 72
- Phillips, L. D., 366, 384,
395
- Pillai, K. C. S., 225, 226,
229
- Policello, G. E. II, 111, 125
- Pollock, S. M., 28, 33
- Pradhan, M., 316, 317
- Preston, E., J., 112, 125
- Proschan, F., 272, 274, 275,
278, 279, 280, 282,
283, 284, 285, 286,
288, 310, 314, 315, 317,
319, 320, 322, 325, 327,
333, 343, 347, 350, 359,
360
- Puri, P. S., 336, 347
- Quandt, R. E., 112, 114, 121,
123, 125, 132,
- Quenouille, M. H., 62, 73
- Quinzi, A. J., 343, 347
- Raiffa, H., 366, 385
- Ramsey, J. B., 112, 114, 121,
123, 125
- Rao, B. R., 77, 85
- Rao, C. R., 112, 125, 224,
226, 230
- Rao, J. S., 323, 325
- Rao, M. M., 275, 279, 287
- Ratnaparkhi, M. V., 331, 334
- Rayment, P. R., 112, 125
- Reid, W. T., 27, 32
- Révész, P., 143, 155
- Reynolds, D. S., 323, 325
- Ricci, A., 134, 135, 139
- Ringer, L. J., 13, 22
- Robertson, C. A., 112, 124,
125
- Robinson, R., 405, 414
- Rose, D. M., 340, 347
- Rosenblatt, M., 141, 147,
149, 152, 154, 155, 169,
180, 352, 361
- Ross, S. M., 284, 288, 350,
361
- Roussas, G. G., 62, 72
- Roy, S. N., 7, 23, 157, 158,
159, 168
- Rozanov, Yu. A., 44, 49, 353,
360
- Rubin, D. B., 103, 108, 119,
124
- Sagrasta, S. N., 290, 308
- Sampford, M. R., 77, 85
- Sanathanan, L. P., 77, 78, 80,
81, 84, 85
- Sarkadi, K., 4, 23, 56
- Sathe, Y. S., 316, 317
- Sauerders, R., 319
- Savage, I. R., 323, 325
- Savits, T. H., 271, 280, 281,
282, 283, 284, 285, 350,
360
- Schafer, R. E., 272, 287
- Scheaffer, R. L., 330, 334
- Scheuer, E. M., 14, 23
- Schever, E. M., 27, 28, 32
- Sen, A. K., 182, 183, 190,
191
- Sen, P. K., 43, 49
- Sethuraman, J., 283, 284, 286,
323, 325
- Schwenk, A., 405, 414
- Shah, S. P., 316, 317
- Shaked, M., 283, 287, 350,
360
- Shannon, C. E., 368, 385

- Shapiro, S.S., 1, 3, 4, 5, 6,
7, 8, 15, 17, 19, 23, 25,
56
- Shaw, L., 279, 287
- Shen, S. M., 378, 384
- Shenton, L. R., 12, 22, 137,
139, 231, 232, 240
- Shimizu, R., 59, 60, 73
- Shirahata, S., 172, 180
- Sibuya, M., 59, 60, 73
- Sidak, Z., 170, 180
- Siddiqui, M. M., 327, 330,
333
- Siegel, S., 366, 384
- Silverman, E., 27, 32
- Simar, L., 97, 99, 105, 109
- Simoni, de, S., 308
- Singh, M., 203, 212, 214
- Singpurwalla, N. D., 272, 282,
283, 286, 287
- Sinha, B. K., 112, 123
- Siotani, M., 207, 214
- Smith, W. L., 79, 85
- Solov'yev, A. D., 13, 22
- Spurrer, J. D., 280, 288
- Srivastava, M. S., 181, 182,
183, 190, 191, 226, 230
- Stacho, L., 55, 146, 155
- Stella, A. L., 350, 360
- Stephens, M. A., 4, 9, 10,
11, 23, 24, 56
- Steutel, F. W., 63, 73
- Stuart, A., 17, 22
- Styen, H. S., 290, 308
- Subbaiah, P., 157, 194, 202
- Subrahmanian, K., 66, 73
- Sullo, P., 274, 275, 279, 288
- Suppes, P., 366, 384
- Tadikamalla, P. R., 231, 240
- Tan, W. Y., 125
- Taqqu, M. S., 354, 359, 361
- Taylor, B. J. R., 112, 125
- Taylor, T. R., 366, 367, 369,
372, 384, 385
- Teicher, H., 96, 109
- Thompson, W. A., 280, 287,
291, 307
- Tietjen, G. L., 11, 22, 204,
214
- Tiku, M. L., 203, 204, 205,
207, 211, 212, 214, 215,
216
- Trivedi, M. C., 158, 167, 219,
- Tsiatis, A., 340, 347
- Tubbs, J. D., 112, 125
- Tukey, J. W., 259, 269, 270
- Tukey, W. L., 79, 85
- Tusnády, G., 143, 144, 155,
156
- Tyan, S. G., 279, 287
- Tyurin, Yu. N., 147, 155
- Uppuluri, V. R. R., 277, 279,
287, 288
- Uven, van M. J., 290, 308
- Van Ness, J. W., 356, 361
- van Uven, M. J., 242, 257
- Vasicek, O., 12, 24
- Vianelli, S., 137, 139
- Voss, R. F., 359, 361
- Wald, A., 222, 229
- Walker, H. F., 112, 125
- Walkup, D. W., 350, 360
- Waller, J. H., 77, 85
- Wallis, J. R., 355, 361
- Wasserman, S., 411, 414
- Watson, D., 80, 81, 82, 85
- Watson, G. S., 36, 49
- Watson, G. S., 328, 334
- Weier, D. R., 169, 172, 173,
179, 180, 280, 288
- Weinman, 279
- Weinrich, M. C., 28, 33
- Weisberg, S., 4, 24
- Weiss, G. A., 328, 334
- Wells, W. T., 328, 334
- West, S. A., 384
- White, J. E., 242, 257
- Wicksell, 275
- Wilk, M. B., 3, 8, 15, 17,
19, 23
- Wilks, S. S., 89, 94, 158,
168, 203, 204, 207,
124
- Williams, C. B., 57, 72
- Wittes, J. T., 77, 86
- Wolfe, J. H., 120, 123, 125

Wolfowitz, J., 97, 108

Wright, E., 414

Yang, G. L., 93, 94

Yoshimura, I., 59, 60,

73

Zachs, S., 182, 191

Zangwill, W. I., 89, 94

Zellner, A., 27, 33

Zhurbenko, I., 353, 361

Zidek, J. V., 313, 317

Zincenko, N. M., 152, 156

SUBJECT INDEX

- approximations, 231
- asymptotically distribution-free test, 51
- asumptotic expansions, 75
- bivariate distribution, 241
- bivariate exponential, 169
- bivariate logarithmic series distribution, 57
- censored samples, 1, 203
- characteristic function, 271
- characterization, 289
- characterizations, 327
- chi-square statistics, 25
- chi-square tests, 35
- chi-squared goodness-of-fit, 57
- coherent system, 309
- combination of tests, 157
- competing risks, 335
- complementary risks, 335
- component life, 309
- consistent tests, 193
- convexity, 387
- Cramér-von Mises type tests of independence, 141
- degree of uncertainty, 363
- dependence concepts, 349
- dependent model, 309
- detecting a change point, 181
- digraphs, 397
- discrete distributions, 259
- distributions, 397
- distributions of minimum and maximum, 335
- distribution of test statistics, 181
- distribution tables and rates of convergence for both, 141
- distributional tests, 1
- EDF tests, 1
- empirical processes, 141
- estimating n , 75
- estimation, 127
- evaluation of distribution functions, 241
- exact slopes, 157
- exponential distribution, 289
- exponential families, 87
- exponential family, 95
- extreme value distribution, 309
- failure rate, 319, 327
- families of statistical distributions, 51
- feature selection discrepancy, 363
- fractional Brownian motion, 349
- gaussian mixtures, 111
- geometrical probability, 387
- gestation model, 271
- goodness-of-fit tests, 35, 51
- graphs, 397
- hazard rate, 289, 309
- hypothesis testing, 193
- identifiability, 335
- independence between two sets of variates, 219
- inequalities, 309
- inference discrepancy, 363
- inferential tasks, 363
- information gain index, 363
- invariance principles, 141
- isoprobability contour, 289
- Johnson's system, 231
- Kendall's tau, 169
- least squares, 231
- life testing, 169
- likelihood ratio test, 25
- likelihood ratio test and estimate, 181
- linear processes, 35
- locally most powerful rank test, 169

- logarithmic transform, 289
- logarithmic transformation, 133
- logistic normal distributions, 363
- maximum likelihood, 87, 111
- maximum likelihood estimators, 25,
- maximum likelihood estimators, 95
- mean residual life function, 327
- mixing, 349
- mixing distributions, 95
- mixture, 327,
- mixtures, 95
- mixtures of distributions, 111
- modes of genesis, 57
- modified maximum likelihood estimators, 203
- moments, 231
- multivariate Cramér-von Mises statistic, 141
- multivariate distribution, 309
- multivariate distributions, 289
- multivariate exponential distribution, 271
- multivariate exponential extensions, 271
- multivariate gamma distribution, 271
- multivariate IFR distributions, 271
- multivariate IFRA, 271
- multivariate logarithmic series distribution, 57
- multivariate mean, 181
- multivariate NBU distributions, 271
- multivariate NBUE distributions, 271
- multivariate outliers, 203
- multivariate Weibull distribution, 271
- noniterative estimation, 87
- normal mixtures, 111
- normative models, 363
- nuisance parameters, 51
- numerical approximation, 127
- objective grouping, 57
- parameter estimation, 133
- Pareto distribution, 127
- Pearson system, 241
- performance analysis, 363
- Pitman relative efficiency, 169
- Poisson process, 319
- power-law spectra, 349
- power study, 169
- power transform, 289
- probabilistic data, 363
- quadrature, 241
- random secants, 387
- rational fraction, 231
- reliability, 335
- renewal process, 327
- robust estimators, 203
- sample size estimation, 75
- second order efficiency, 75
- self-similarity, 349
- series and parallel systems, 335
- shock models, 271
- shock model, 319
- Spearman's rho, 169
- sphericity, 219
- statistics, 397
- stepdown procedure, 157
- stochastic expansion, 25
- stochastic failure rate, 319
- stochastic processes, 349
- strictly stationary and strong mixing processes, 35
- subject inference, 363
- survival analysis, 327
- system life, 309

t-distribution, 193
test for lognormality, 133
test of independence, 169
threshold model, 271
transformation, 231
transformations, 259
truncated samples, 75
t-tests, 193

unbiased tests, 193

weak convergence, 319
Weibull, 279, 280
Weibull distribution, 309, 319
weighed Cramér-von Mises
test, 51
W tests, 1

